

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2000, 20:18:44 ; Search time 360.16 Seconds
(without alignments)
-3498.721 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419
Perfect score: 415
Sequence: 1 ACCGGGATTCAGTGTCTCC.....AAAATTCGTATGCTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_cm.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_v1.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_lo.*
- 22: em_lo.*
- 23: em_lo.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pi.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_y1.*
- 32: gb_hcg1.*
- 33: gb_hcg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: gb_in3.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_hcg3.*
- 42: gb_hcg4.*
- 43: gb_hcg5.*
- 44: gb_hcg6.*
- 45: gb_hcg7.*
- 46: gb_hcg1.*
- 47: em_hcg2.*
- 48: em_hcg3.*
- 49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	221	53.3	47323	11	AC005937	AC005937 Homo sapi
2	219.4	52.9	192650	10	AB023048	AB023048 Homo sapi
3	219.4	52.9	200000	10	AF000511	AF000511 Homo sapi
	42.2	10.2	7218	5	I66494	I66494 Sequence 14
C 5	37.2	9.0	185548	45	AC012467	AC012467 Homo sapi
C 6	36.4	8.8	118276	11	AC004148	AC004148 Homo sapi
C 7	36.2	8.7	65870	12	AC005807	AC005807 Mus muscu
C 8	35.4	8.5	207984	33	AC006329	AC006329 Homo sapi
9	35	8.4	207092	44	AC013564	AC013564 Homo sapi
10	34.8	8.4	148231	42	AC011995	AC011995 Homo sapi
C 11	34.6	8.3	65673	42	AC012184	AC012184 Homo sapi
C 12	34.4	8.3	77322	10	HS0144C9	HS0144C9 Human DNA
C 13	34.2	8.2	3626	12	D86631	D86631 Mus sp. DNA
14	34	8.2	3676	11	HSJ858B16	HSJ858B16 Human DNA
C 15	33.8	8.1	62181	44	AC016298	AC016298 Homo sapi
C 16	33.6	8.1	128379	11	HSJ773A18	HSJ773A18 Human DNA
C 17	33.6	8.1	25065	43	AC014817	AC014817 Drosophil
C 18	33.4	8.0	112309	11	AC003025	AC003025 Human Chr
C 19	33.4	8.0	185035	11	AC004770	AC004770 Homo sapi
C 20	33.4	8.0	196080	40	AC004228	AC004228 Homo sapi
C 21	33.4	8.0	142032	40	AF139813	AF139813 Homo sapi
C 22	33.2	8.0	187000	10	AF000059	AF000059 Aeropyrum
C 23	33.2	8.0	103182	43	HS158C21	HS158C21 Human DNA
C 24	33.2	8.0	116218	11	AC005500	AC005500 Homo sapi
C 25	33	8.0	124700	35	AC005509	AC005509 Homo sapi
C 26	33	8.0	124700	35	AC005509	AC005509 Homo sapi
C 27	32.8	7.9	30452	10	AB014077	AB014077 Drosophil
C 28	32.8	7.9	100000	10	AF000513	AF000513 Homo sapi
C 29	32.8	7.9	43600	11	AC004209	AC004209 Homo sapi
C 30	32.8	7.9	159515	12	AC002327	AC002327 Mus muscu
C 31	32.8	7.9	19508	34	CFC14C10	CFC14C10 Caenorhabdi
C 32	32.6	7.9	5280	9	AB011114	AB011114 Homo sapi
C 33	32.6	7.9	172048	10	HS179N16	HS179N16 Homo sapien
C 34	32.6	7.9	156197	11	HS97K10	HS97K10 Human DNA s
C 35	32.4	7.8	2255	8	AF052391	AF052391 Trichophy
C 36	32.4	7.8	2458	8	AF052391	AF052391 Trichophy
C 37	32.4	7.8	184427	16	EHV020824	EHV020824 Equine herp
C 38	32.2	7.8	13882	2	AE001157	AE001157 Borrelia
C 39	32.2	7.8	2122	3	PIGATA	PIGATA Porcine cho
C 40	32.2	7.8	2122	3	PIGATA	PIGATA Porcine cho
C 41	32.2	7.8	1277	7	EMEATGSS9	EMEATGSS9 M30144 A.nidulans
C 42	32.2	7.8	123203	11	AC003036	AC003036 Homo sapi
C 43	32.2	7.8	167228	40	AC005552	AC005552 Homo sapi
C 44	32	7.7	37036	1	MSGY151	AD000018 Mycobacte
C 45	32	7.7	32514	1	MTCY130	273902 Mycobacteri

ALIGNMENTS

RESULT	1	AC005937	47323 bp	DNA	PRI	05-NOV-1998
LOCUS		Homo sapiens clone UMG	370M23.002	from 6p21, complete sequence.		
DEFINITION		AC005937				
ACCESSION		AC005937				
VERSION		AC005937.1	GI:3845393			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia, Eutheria; Primates; Catarrhini; Hominidae; Homo				
		1 (bases 1 to 47323)				

QY 255 CTCACACCATGTTGCTGCAACACATGACAGCCATTGAAGCCTGTGTCCTTTGGCCCC 314
 Db 35747 CTCACACCATGTTGCTGCAACACATGACAGCCATTGAAGCCTGTGTCCTTTGGCCCC 35806
 QY 315 GGGCTTTGGCCGGGGATGCGAGGAGGCCCGCCGACCTGTCTTTTCAGCAGGCCCCCA 374
 Db 35807 GGGCTTTGGCCGGGGATGCGAGGAGGCCCGCCGACCTGTCTTTTCAGCAGGCCCCCA 35866
 QY 375 CCTCCTGAGTGCGCAATAAATAAATTCGGTATGCTGAATT 415
 Db 35867 CCTCCTGAGTGCGCAATAAATAAATTCGGTATGCTGAATT 35907
 RESULT 2
 LOCUS AB023048 192650 bp DNA PRI 20-NOV-1999
 DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
 clone:53L9, complete sequence.
 ACCESSION AB023048
 VERSION HTG.
 KEYWORDS
 SOURCE Homo sapiens cell_line:978SK DNA, clone:53L9.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Kikawa,T., Tamiya,G., Oka,A., Takishima,N., Yanagata,T.,
 Kikawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
 Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
 Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
 Bahram,S. and Inoko,H.
 TITLE Molecular dynamics of MHC genesis unraveled by sequence analysis of
 the 1,796,938-bp HLA class I region
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
 MEDLINE 20027539
 REFERENCE 2 (bases 1 to 192650)
 AUTHORS Shina,T. and Takishima,N.
 TITLE Direct Submission
 JOURNAL Submitted (25-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
 Shina, Tokai University School of Medicine, Department of
 Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
 Japan (E-mail:tsuhina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
 Fax:81-463-94-8884)
 FEATURES
 source Location/Qualifiers
 1. .192650
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /cell_line="978SK"
 /chromosome="6"
 /clone="53L9"
 /map="6p21.3"
 BASE COUNT 49862 a 44743 c 45833 g 52212 t
 ORIGIN
 Query Match 52.9%; Score 219.4; DB 10; Length 192650;
 Best Local Similarity 99.5%; Pred. No. 4.9e-55;
 Matches 220; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 195 AGCTGTGCTCTCTCCTCCATCTCCTTCAGGGACCGCTCACCCTCCACCATGCAAGAT 254
 Db 178201 AGCTGTGCTCTCTCCTCCATCTCCTTCAGGGACCGCTCACCCTCCACCATGCAAGAT 178260
 QY 255 CTCACACCATGTTGCTGCAACACATGACAGCCATTGAAGCCTGTGTCCTTTGGCCCC 314
 Db 178261 CTCACACCATGTTGCTGCAACACATGACAGCCATTGAAGCCTGTGTCCTTTGGCCCC 178320
 QY 315 GGGCTTTGGCCGGGGATGCGAGGAGGCCCGCCGACCTGTCTTTTCAGCAGGCCCCCA 374
 Db 178221 GGGCTTTGGCCGGGGATGCGAGGAGGCCCGCCGACCTGTCTTTTCAGCAGGCCCCCA 178380
 QY 375 CCTCCTGAGTGCGCAATAAATAAATTCGGTATGCTGAATT 415

Db 178381 CCTCCTGAGCGGCATTAATAAATTCGGTATGCTGAATT 178421
 RESULT 3
 LOCUS AP000511 200000 bp DNA PRI 28-SEP-1999
 DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
 section 10/20.
 ACCESSION AP000511
 VERSION GI:5926698
 KEYWORDS
 SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (sites)
 AUTHORS Shina,S., Tamiya,G., Oka,A. and Inoko,H.
 TITLE Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
 JOURNAL Published Only in DataBase (1999) in press
 REFERENCE 2 (bases 1 to 200000)
 AUTHORS Hirakawa,M., Yanaguchi,H., Imai,K. and Shimada,J.
 TITLE Direct Submission
 JOURNAL Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
 Hirakawa, Japan Science and Technology Corporation (JST), Advanced
 Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
 Japan (E-mail:mika@tokyo.jst.go.jp,
 URL:http://www-alis.tokyo.jst.go.jp,
 Fax:81-3-5214-8470)
 COMMENT This sequence is conducted by Tokai University as a JST sequencing
 Team.
 Principal Investigator: Hidetoshi Inoko Ph.D
 Phone:+81-463-93-1121, Fax:+81-463-94-8884,
 The sequence is submitted by Human Genome Sequencing in ALIS
 project of JST
 Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
 For further information about this sequences, please visit our
 sequence archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
 html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
 FEATURES
 source Location/Qualifiers
 1. .200000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p21.3"
 108774. .108864
 /standard_name="D6S2088"
 /note="SHGC-12985;The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:735268"
 115648. .115906
 /standard_name="D6S1898"
 /note="WI-9418;The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:678272"
 complement(115665..115896)
 /note="cda0vhl0;The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:443382"
 142030. .142159
 /standard_name="D6S2203"
 /note="SHGC-16870;The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:741215"
 complement(142085..142322)
 /note="RH18132;The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:4573021"
 complement(142479..142740)
 /standard_name="D6S1851"
 /note="SHGC-10808;The location is between each flanking
 site of PCR primers."

[illegible]

repeat_region	complement(3391..3401). /rpt family="AluSg"
repeat_region	complement(3402..3428) /rpt family="TTAAAn"
repeat_region	complement(3429..3677) /rpt family="AluSg"
repeat_region	3714..3884 /rpt family="FRAM"
repeat_region	4143..4219 /rpt family="L2"
repeat_region	complement(5060..5364) /rpt family="AluSg"
repeat_region	6907..7148 /rpt family="L2"
repeat_region	7161..7426 /rpt family="L2"
repeat_region	7508..7647 /rpt family="L2"
repeat_region	complement(9623..9918) /rpt family="AluSx"
repeat_region	10006..10304 /rpt family="AluSg1"
repeat_region	10312..10478 /rpt family="AluSg"
repeat_region	10508..10530 /rpt family="At.Rich"
repeat_region	complement(15532..15839) /rpt family="AluSx"
repeat_region	complement(15844..16126) /rpt family="AluSx"
repeat_region	17035..17334 /rpt family="AluSx"
repeat_region	17423..17720 /rpt family="AluSx"
repeat_region	complement(18239..18513) /rpt family="AluJo"
repeat_region	18587..18814 /rpt family="MIR"
repeat_region	complement(20103..20260) /rpt family="LIME3A"
repeat_region	complement(20340..20494) /rpt family="AluJb"
repeat_region	complement(20495..20793) /rpt family="AluSx"
repeat_region	complement(20794..20934) /rpt family="AluJb"
repeat_region	complement(20970..21272) /rpt family="AluSg"
repeat_region	complement(21467..21955) /rpt family="L2"
repeat_region	complement(21966..22276) /rpt family="L2"
repeat_region	complement(22808..22989) /rpt family="MBR46A"
repeat_region	complement(23074..23356) /rpt family="AluSx"
repeat_region	complement(23361..23466) /rpt family="LIMB8"
repeat_region	23467..23759 /rpt family="AluSg"
repeat_region	complement(23760..24596) /rpt family="LIMB8"
repeat_region	24597..24889 /rpt family="AluJb"
repeat_region	24897..25249 /rpt family="MTID"
repeat_region	complement(25255..25400) /rpt family="LIM4"
repeat_region	complement(25404..25427) /rpt family="L1"
repeat_region	complement(25428..25741) /rpt family="AluSx"
repeat_region	complement(25742..25828) /rpt family="AluSx"

http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This 'working draft' quality sequence may consist of several contigs from automated sequence assembly concatenated together. No attempt has been made to order or orient the contigs relative to one another correctly before concatenating. At each location in the sequence where contigs have been joined, several consecutive Ns may have been inserted.

The attached annotation was produced using a purely automated procedure.

The location of this clone is unknown.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

FEATURES

Source

```
1. 148231
   Location/Qualifiers
    organism="Homo sapiens"
    db_xref="taxon:9606"
    chromosome="unknown"
    clone="NH016321"
    clone_lib="unknown"
    1840..1961
    /rpt_family="MIR"
    2039..2122
    /rpt_family="MIR"
    2196..2601
    /rpt_family="L1"
    3563..3714
    /rpt_family="MIR"
    3947..4338
    /rpt_family="L2"
    5774..5951
    /rpt_family="(TG)n"
    8803..8851
    /rpt_family="MIR"
    9022..9213
    /rpt_family="MIR"
    9402..10001
    /rpt_family="MIR"
    10353..11049
    /rpt_family="MER4-group"
    /note="pseudogene similar to A28996 (PID:g91208)
    11110..11202
    /rpt_family="MaLR"
    11951..12041
    /rpt_family="MIR"
    12532..12660
    /rpt_family="MIR"
    12697..12719
    /rpt_family="(TTTA)n"
    12720..13002
    /rpt_family="Alu"
    14466..14509
    /rpt_family="T-rich"
    14511..15076
    /rpt_family="L1"
    15804..15829
    /rpt_family="(TTCA)n"
    16282..16416
    /rpt_family="Retroviral"
    16750..16798
    /rpt_family="(TA)n"
    16798..16842
    /rpt_family="(CA)n"
    19964..19986
    /rpt_family="AT-rich"
    20606..20663
    /rpt_family="Retroviral"
    20665..20723
```

```
repeat_region      /rpt_family="(TA)n"
21232..21253
repeat_region      /rpt_family="AT-rich"
21415..21876
repeat_region      /rpt_family="L2"
22093..22165
repeat_region      /rpt_family="MIR"
23940..24094
repeat_region      /rpt_family="(TG)n"
24163..24325
repeat_region      /rpt_family="(TATATG)n"
24304..24451
repeat_region      /rpt_family="(TG)n"
24462..24524
repeat_region      /rpt_family="MER4-group"
24525..25055
repeat_region      /rpt_family="(TG)n"
25328..25695
repeat_region      /rpt_family="MaLR"
25931..26026
repeat_region      /rpt_family="L2"
27904..27962
repeat_region      /rpt_family="GA-rich"
28278..28730
repeat_region      /rpt_family="L2"
28917..29288
repeat_region      /rpt_family="MaLR"
30027..30128
repeat_region      /rpt_family="MIR"
32576..32706
repeat_region      /rpt_family="MIR"
33879..34233
repeat_region      /rpt_family="MER21-group"
34234..34593
repeat_region      /rpt_family="MaLR"
34594..35003
repeat_region      /rpt_family="MER21-group"
35566..36099
repeat_region      /rpt_family="MIR"
36805..36841
repeat_region      /rpt_family="Alu"
37195..37254
exon               38143..38283
exon               38442..38696
repeat_region      /rpt_family="G-rich"
39567..39779
misc_feature        /note="pseudogene similar to P35085 (PID:g461699) CALCION
BINDING PROTEIN"
40070..40288
repeat_region      42643..42960
misc_feature        /rpt_family="MER2-type"
44800..45117
repeat_region      /note="pseudogene similar to PID:g1070603 collagen alpha
1(III) chain precursor - human"
45696..45752
repeat_region      /rpt_family="L2"
46140..46559
repeat_region      /rpt_family="L2"
46918..47149
repeat_region      /rpt_family="MaLR"
47503..47571
repeat_region      /rpt_family="MIR"
49216..49583
repeat_region      /rpt_family="MaLR"
49866..50010
repeat_region      /rpt_family="MIR"
51046..51183
repeat_region      /rpt_family="L2"
51453..51806
repeat_region      /rpt_family="MaLR"
51807..52111
repeat_region      /rpt_family="Alu"
52112..52185
```

```
repeat_region /rpt_family="MaLR"
52475..52955
repeat_region /rpt_family="L2"
53870..54109
repeat_region /rpt_family="MaLR"
54110..54357
repeat_region /rpt_family="Alu"
54358..54418
repeat_region /rpt_family="MaLR"
54419..54720
repeat_region /rpt_family="Alu"
54721..54803
repeat_region /rpt_family="MaLR"
55004..55659
repeat_region /rpt_family="L1"
55769..55940
repeat_region /rpt_family="MIR"
55944..56399
repeat_region /rpt_family="MaLR"
57034..57065
repeat_region /rpt_family="AT-rich"
57430..57728
repeat_region /rpt_family="Alu"

Query Match 8.4% Score 34.8; DB 42; Length 148231;
Best Local Similarity 43.5%; Pred. No. 6.6;
Matches 108; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 141 AGTCTTTCTGACAAATCCCTATGAGTCCAGCTTCTCGAATTCCTTGAAGCTCT 200
Db 68150 AATATCTGTTCAAAATCCCTGTTGACCAAGATATGAGGAGCATGGAGATTCNN 68209
QY 201 GCCTCCTCCTCCATCTCCCTTCAGGACAGCGTCACCTCCAGCATGCAAGATCTCAC 260
Db 68210 NNNNNNNNNNNNTGGAGCTGTATCCCTTTTCCCAACCAACAGACAGCTCTAC 68269
QY 261 ACCATGTTCTCAACATGACAGCATGAGCTGTCTCTTGGCCCGGGCTT 320
Db 68270 TGTAGTAGTACAGCTGCCCTCTCTGCCCCGAGCGGTGTCTCCAGTGCCACGGTT 68329
QY 321 TTGGCCCGGGATGACAGGAGGAGCGCCCGACCTCTCTTTCACAGCGCCCGCCCTCC 380
Db 68330 CTTGCCCTGGGATACCTCTGCCCCAGCGCCCTGAGCTTCTTGTGCTGCTCCCTCC 68389
QY 381 TGAGTGGC 388
Db 68390 TGTGGAGC 68397

RESULT 11
AC012184/c 65673 bp DNA HTG 21-OCT-1999
LOCUS Homo sapiens chromosome 16 clone RPC1-11_529K1, *** SEQUENCING IN
DEFINITION PROGRESS **, 40 unordered pieces.
ACCESSION AC012184
VERSION AC012184.1 GI:6091684
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 65673)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65673)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (21-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
www.jgi.doe.gov.
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 40 contigs. The true order of the pieces
```

```
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 760: contig of 760 bp in length
gap of unknown length
761 1560: contig of 800 bp in length
gap of unknown length
1561 2303: contig of 743 bp in length
gap of unknown length
2304 3255: contig of 952 bp in length
gap of unknown length
3256 4246: contig of 991 bp in length
gap of unknown length
4247 5603: contig of 1357 bp in length
gap of unknown length
5604 6762: contig of 1159 bp in length
gap of unknown length
6763 7004: contig of 242 bp in length
gap of unknown length
7005 8046: contig of 1042 bp in length
gap of unknown length
8047 9309: contig of 1263 bp in length
gap of unknown length
9310 10988: contig of 1679 bp in length
gap of unknown length
10989 12037: contig of 1049 bp in length
gap of unknown length
12038 13400: contig of 1363 bp in length
gap of unknown length
13401 14099: contig of 699 bp in length
gap of unknown length
14100 15399: contig of 1300 bp in length
gap of unknown length
15400 16772: contig of 1373 bp in length
gap of unknown length
16773 17868: contig of 1096 bp in length
gap of unknown length
17869 18643: contig of 775 bp in length
gap of unknown length
18644 19788: contig of 1145 bp in length
gap of unknown length
19789 21845: contig of 2057 bp in length
gap of unknown length
21846 23223: contig of 1378 bp in length
gap of unknown length
23224 25010: contig of 1787 bp in length
gap of unknown length
25011 25466: contig of 456 bp in length
gap of unknown length
25467 27098: contig of 1632 bp in length
gap of unknown length
27099 28274: contig of 1176 bp in length
gap of unknown length
28275 29748: contig of 1474 bp in length
gap of unknown length
29749 31363: contig of 1615 bp in length
gap of unknown length
31364 33150: contig of 1787 bp in length
gap of unknown length
33151 36067: contig of 2917 bp in length
gap of unknown length
36068 37581: contig of 1514 bp in length
gap of unknown length
37582 39288: contig of 1707 bp in length
gap of unknown length
39289 42139: contig of 2851 bp in length
gap of unknown length
42140 44809: contig of 2670 bp in length
gap of unknown length
44810 47022: contig of 2213 bp in length
```

```

* 47023 51524: contig of 4502 bp in length
* 51525 53246: contig of 1722 bp in length
* 53247 56192: contig of 2946 bp in length
* 56193 58896: contig of 2704 bp in length
* 58897 61546: contig of 2650 bp in length
* 61547 65673: contig of 4127 bp in length.
FEATURES             Location/Qualifiers
     source           1..65673
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RPCI-11.529K1"
BASE COUNT      16932 a 16632 c 16336 g 15686 t      87 others
ORIGIN

```

FEATURES

source

This sequence is the entire insert of clone RPI-144C9.

Location/Qualifiers

1..77322

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="1"

/map="p34.3-36.11"

/clone="RPI-144C9"

/clone_lib="RPCI-1"

complement(1710..2227)

/note="match: GSS: Em:AQ355665"

complement(1780..2227)

/note="match: GSS: Em:AQ355618"

2858..3151

/note="match: GSS: Em:AQ627537.1"

complement(11439..11580)

/note="match: GSS: Em:AQ031787"

complement(24560..25030)

/note="match: GSS: Em:AQ307331"

complement(24637..25031)

/note="match: GSS: Em:AQ234890"

27999..28360

/note="match: STS: Em:G14651"

29391..29649

/note="match: STS: Em:G06203"

complement(37609..37922)

/note="match: GSS: Em:AQ114724"

47849..48201

/note="match: GSS: Em:AQ553574.1"

63243..63496

/note="match: GSS: Em:AQ142826"

BASE COUNT 21006 a 18901 c 18290 g 19125 t

ORIGIN

Query Match

Best Local Similarity 8.3%; Score 34.4; DB 10; Length 77322;

Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 14 TGTCTCTCCATCCAGGAGCGAGTGGCCATGATGGGCTGTGGGCTGCGCCCTTGTCCTC 73

Db 31300 TGTTCCTCCCTCCAGGATCCCTTTGGTGATGTTTCCAGGATGCACCACCAC 31241

QY 74 CTCTTGACCTCTCTGCGAGCTCA 97

Db 31240 CTCTAGATACCTTCAGGCAACACA 31217

RESULT 13

LOCUS

D86631/c

DEFINITION

Mus sp. DNA for oxytocin receptor, partial cds.

ACCESSION

D86631

VERSION

D86631.1

KEYWORDS

oxytocin receptor

SOURCE

Mus sp. (Strain:129SV) ES cell DNA.

ORGANISM

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 3626)

AUTHORS

Kubota,Y.

TITLE

Direct Submission

JOURNAL

Submitted (26-JUL-1996) to the DDBJ/EMBL/GenBank databases. Yasue

```

* 47023 51524: contig of 4502 bp in length
* 51525 53246: contig of 1722 bp in length
* 53247 56192: contig of 2946 bp in length
* 56193 58896: contig of 2704 bp in length
* 58897 61546: contig of 2650 bp in length
* 61547 65673: contig of 4127 bp in length.
FEATURES             Location/Qualifiers
     source           1..65673
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RPCI-11.529K1"
BASE COUNT      16932 a 16632 c 16336 g 15686 t      87 others
ORIGIN

```

```

Query Match
Best Local Similarity 8.3%; Score 34.6; DB 42; Length 65673;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

```

QY 190 TGAAGACTCTGCTCTCCATCCCTTCAGGAGCGAGTGGCCATGATGGGCTGTGGGCTGCGCCCTTGTCCTTC 249
Db 39815 TGAAGAACTCTCCATCCCTTCAGGAGCGAGTGGCCATGATGGGCTGTGGGCTGCGCCCTTGTCCTTC 249

```

```

QY 250 AAGATCTACACGAGTGTCTGCACACATGACAGCGCAATGAAGCCTGTGCTCTCTT 309
Db 39755 TAACTTCAAGCGAATTCCTGTGATGCGCGCTCTCTTTTGAGAATTAATCTTT 3696

```

```

QY 310 GCCCGGGC 318
Db 39695 GATCTGGC 39687

```

RESULT 12

HSDJ144C9/c

LOCUS

Human DNA sequence from clone RPI-144C9 on chromosome 1p34.3-36.11,

complete sequence.

AL096774

VERSION

AL096774.9

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

Hall,R.

Submitted (22-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonesrequest@sanger.ac.uk

On Nov 22, 1999 this sequence version replaced gi:6138775.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The following abbreviations are used to associate primary accession

numbers given in the feature table with their source databases:

Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information


```

/feature="dJ858B16.1"
/note="match: cDNAs: Em:AB011114"
/evidence-not_experimental
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
273..16218
/feature="dJ858B16.1"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12717..12948,1414..14546,
14669..14747,15784..15895,15983..16087)
/feature="dJ858B16.1"
/note="different from published protein but supported by
GENES and GENSCAN"
/codon_start=1
/evidence-not_experimental
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
/protein_id="CAB56392.1"
/db_xref="GI:5921489"
/translators="OAILHAERQLLYRWFVWQQAARHQQEWMQVACAHHRGRLL
KACFLRESAQLRTERGRVRAEFHMAQLLRWAWSONRECLALRGAROKLMRAD
LHQHSLVLRALQAVTYQGRVRSILREVAAESQHNQLLRGALREKENTMARVDE
AKKTFQASTHYRTICSKVLVQWREAVSYQMYRQEDCAIWEAQKVLDRGLRTWFO
RWDCSRSQAQQLQARQVORHRRQLLEGLARWTKHLLQVRLKLLHROSTQLAQ
RLSTCFRQWRQQLAARQEQATVRLAFWAFSLQAKVWATLAFVLERRKKLRLQ
WALQAYQGLQEGATRLLRFAASMKASRQQLAQOQVQAHSIHLRAVRCATLWKQ
VLRGGGQPLAAIAPSRKVTFFEGPLNRIAGAGDGTLEKRPQASRPLGALGLAA
EPHALPQKPOEHLGMAQPAAPSLTRFFLAEPATLVPHSLPGLASSAPGPKQP
PTASTGPPELLLLPLSEFPCGAAPARVSPQATPRDKPPVPFSSLASVDPDHLPLPGD
PSNTRAGPLSTAGSDLEAELEIQOQLHYOTTQNLWSCRQASRLRWLELNRE
EPGDEDEQQVQKLEQVEMQIOLLAELAQRPQIGACVARIQALQALC"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11528)
/feature="dJ858B16.1"
/note="Original published form; could be artefactual as it
is result of a large intron read-through; match: proteins:
Tr:060289"
/codon_start=1
/evidence-not_experimental
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
/protein_id="CAB56393.1"
/db_xref="GI:5921490"
/translators="OAILHAERQLLYRWFVWQQAARHQQEWMQVACAHHRGRLL
KACFLWRESAQLRTERGRVRAEFHMAQLLRWAWSONRECLALRGAROKLMRAD
LHQHSLVLRALQAVTYQGRVRSILREVAAESQHNQLLRGALREKENTMARVDR
AKKTFQASTHYRTICSKVLVQWREAVSYQMYRQEDCAIWEAQKVLDRGLRTWFO
RWDCSRSQAQQLQARQVORHRRQLLEGLARWTKHLLQVRLKLLHROSTQLAQ
RLSTCFRQWRQQLAARQEQATVRLAFWAFSLQAKVWATLAFVLERRKKLRLQ
WALQAYQGLQEGATRLLRFAASMKASRQQLAQOQVQAHSIHLRAVRCATLWKQ
VLRGGGQPLAAIAPSRKVTFFEGPLNRIAGAGDGTLEKRPQASRPLGALGLAA
EPHALEL"
620..913
/note="AluX repeat: matches 1..294 of consensus"
1444..1687
/feature="dJ858B16.1"
/note="match: GSS: Em:AQ545234"
3250..3336
/note="AluJo repeat: matches 2..302 of consensus"
3634..3836
/note="L2 repeat: matches 2497..2701 of consensus"
4356..4646
/note="AluX repeat: matches 21..312 of consensus"
4650..4754
/note="AluSg/x repeat: matches 200..307 of consensus"
4990..5281
/note="AluX repeat: matches 1..292 of consensus"
5806..6231
/note="MLTIC repeat: matches 6..460 of consensus"
6529..6668
/note="FRAM repeat: matches 1..140 of consensus"
6671..6966
/note="AluY repeat: matches 1..295 of consensus"
6985..7081

```

```

repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
repeat_region
polyA_site
mRNA
/note="LTR40a repeat: matches 1..100 of consensus"
7929..8097
/note="MIR repeat: matches 82..255 of consensus"
9766..10068
/note="AluY repeat: matches 1..301 of consensus"
10069..10096
/note="L4 copies 2 mer ta 100% conserved"
10104..10399
/note="AluJb repeat: matches 1..293 of consensus"
12099..12259
/note="MIR repeat: matches 46..208 of consensus"
13063..13779
/note="L2 repeat: matches 1570..2750 of consensus"
15108..15404
/note="AluSg repeat: matches 1..295 of consensus"
complement(16160)
/feature="dJ858B16.2"
complement(join(16160..17505,18222..18382,18665..18811,
19003..19141,19318..19554,21353..21518,23432..23685,
28473..28494))
/feature="dJ858B16.2"
/note="match: cDNAs: Em:M62722 Em:AL050371 Em:AF086277
Em:AF035304 Em:U82163 Em:AB011114; match: ESTs:
Em:AA324939 Em:AA682112 Em:AA166445 Em:AA166440
Em:AA3260704 Em:A1019759 Em:AL039237 Em:AA118158
Em:AV083921 Em:A1232370 Em:A0035550 Em:AV12327
Em:AV17921 Em:AA326672 Em:A0051210 Em:AA298268
Em:AA144920 Em:AA673371 Em:R00603 Em:AA297077 Em:AA237290
Em:AA15772 Em:AA297754 Em:AA297076 Em:AA643609
Em:AI094871 Em:AA298523 Em:AA149582 Em:AA836285
Em:N031353 Em:AA297159 Em:A1191287 Em:AA916775 Em:R51102
Em:A01773 Em:AA521233 Em:AA044074 Em:A1361548 Em:H69648
Em:AA971234 Em:W00857 Em:A1623440 Em:AA553345 Em:H93993
Em:AA046738 Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086
Em:W69536 Em:T10522 Em:A1024124 Em:H71693 Em:AA302951
Em:N98260 Em:N63934 Em:AA215768 Em:R94341 Em:AA177342
Em:R48874 Em:A1342072 Em:W90666 Em:AA255719 Em:A1024249
Em:AA829254 Em:W69460 Em:AA595561 Em:W69544 Em:R77685
Em:AV179744 Em:N64034 Em:AA075750 Em:H38339 Em:AV095355
Em:AV116402 Em:A1425026 Em:AA974403 Em:AA908164
Em:AA088219 Em:AA298551 Em:N56896 Em:AV030271 Em:A1428571
Em:R006913 Em:A1014094 Em:H53175 Em:A1809747 Em:H38743
Em:H62728 Em:W03262 Em:R52264 Em:AA044185 Em:R82229
Em:AA369042 Em:AA939183 Em:A1126154 Em:R54847 Em:AA158770
Em:AA825239 Em:R53399 Em:A1770152 Em:AA594535 Em:AA582394

```

```

Query Match 8.2%; Score 34; DB 11; Length 36676;
Best Local Similarity 54.9%; Pred. No. 10;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
Qy 15 GTCTCTCCATCCAGGACGCGCATGTCGGGTCTGGGTCTGGCTTGTCTCTCC 74
Db 14428 GGCTACTCTAGGATAAGCCCGGTCTCCCTCATCTCTGGCCAGTGTCTCTGACCCCA 14487
Qy 75 TCTTGACCTCTCTGACGTCACATGACAGCGCGCGGTATGACTTTGCAACTGAAC 134
Db 14488 TCTACTCTCTCTGGGACCTTCTAGCCACAGGCTGGGCTTGACTTTCAACTGCAGG 14547
Qy 135 TG 136
Db 14548 TG 14549
RESULT 15
AC016298
LOCUS AC016298 62181 bp DNA HTG 24-NOV-1999
DEFINITION Homo sapiens Clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC016298
VERSION AC016298.1 GI:6467036
KEYWORDS HTG; HTGS-PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; ; ;

```



```

* 40011 40741: contig of 731 bp in length
* gap of unknown length
* 40742 41570: contig of 829 bp in length
* gap of unknown length
* 41571 42298: contig of 728 bp in length
* gap of unknown length
* 42299 42980: contig of 682 bp in length
* gap of unknown length
* 42981 43786: contig of 806 bp in length
* gap of unknown length
* 43787 44584: contig of 798 bp in length
* gap of unknown length
* 44585 45392: contig of 808 bp in length
* gap of unknown length
* 45393 46206: contig of 814 bp in length
* gap of unknown length
* 46207 46991: contig of 785 bp in length
* gap of unknown length
* 46992 47813: contig of 822 bp in length
* gap of unknown length
* 47814 48623: contig of 810 bp in length
* gap of unknown length
* 48624 49437: contig of 814 bp in length
* gap of unknown length
* 49438 50183: contig of 746 bp in length
* gap of unknown length
* 50184 50985: contig of 802 bp in length
* gap of unknown length
* 50986 51800: contig of 815 bp in length
* gap of unknown length
* 51801 52595: contig of 795 bp in length
* gap of unknown length
* 52596 53395: contig of 800 bp in length
* gap of unknown length
* 53396 54205: contig of 810 bp in length
* gap of unknown length
* 54206 54996: contig of 791 bp in length
* gap of unknown length
* 54997 55783: contig of 787 bp in length
* gap of unknown length
* 55784 56587: contig of 804 bp in length
* gap of unknown length
* 56588 57389: contig of 802 bp in length
* gap of unknown length
* 57390 58203: contig of 814 bp in length
* gap of unknown length

Query Match      8.18; Score 33.8; DB 44; Length 62181;
Best Local Similarity 53.48; Pred. No. 12;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 6 GACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCC 65
   |||||
Db 11905 GATTTTAATTCCTTACCCCTGAAGTGGAGTGGTCTCTCTTTGGTGGCTCCCCC 11964
   |||||

QY 66 TTGCTCTCTCTGTGACCTCTCTTGGCAGCTCACATGGAACAGCGCGGTATGACTTGC 125
   |||||
Db 11965 TCTTCTTCTTGGCTCTGGGTGGCAGTCCCTGGAAGAAGTCTCAATCCCTTGC 12024
   |||||

QY 126 AACTGAAGCTGAA 138
   |||||
Db 12025 GCCTCGGGTTGCA 12037

```

Search completed: March 18, 2000, 20:24:08
Job time: 1518 sec

PR 16-APR-1997; US-843652.
 PI (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PA Holtzman D;
 DR WPI: 98-594562/50.
 DR P-PSDB; W79261.
 PT Isolated tumour necrosis factor related proteins - used to develop
 PT products for the diagnosis and treatment of apoptosis-related
 PT disorders, e.g. cancers, autoimmune disorders of neurodegenerative
 PT disorders.
 PS Claim 7; Fig 2; 88pp; English.
 CC This cDNA sequence includes a coding region for a new member of the
 CC human tumour necrosis factor receptor superfamily that is designated
 CC Tango-63e (see W79261). 2 Different forms of Tango-63, i.e.
 CC Tango-63e and Tango-63d (see W62672), were identified in a human
 CC prostate cDNA library through EST sequencing and screening. The
 CC encoded polypeptides are identical with the exception of the
 CC deletion of amino acids 183-211 of Tango-63d (see W79260) in
 CC Tango-63e. The Tango-63 gene was mapped to chromosome 8 between
 CC markers WI-6088 and WI-6563. A plasmid encoding Tango-63e is
 CC deposited as ATCC 98367. The invention encompasses nucleic acid
 CC molecules encoding Tango-63d and -63e, vectors containing these
 CC nucleic acid molecules, cells harboring recombinant DNA encoding
 CC Tango-63d and/or -63e, fusion proteins that include Tango-63d
 CC and/or -63e, transgenic animals that express Tango-63d and/or -63e,
 CC and recombinant knockout animals that fail to express Tango-63d
 CC and/or -63e. Methods are provided for the diagnosis and treatment
 CC of disorders associated with either an abnormally high or an
 CC abnormally low rate of apoptotic cell death. Inhibitors can be
 CC used for treating e.g. cancers, autoimmune disorders (e.g. systemic
 CC lupus erythematosus and immune-mediated glomerulonephritis), and
 CC viral infections (e.g. herpesviruses, poxviruses, and adenoviruses).
 CC Agonists can be used for treating e.g. neurodegenerative diseases,
 CC e.g. Alzheimer's disease, Parkinson's disease, amyotrophic lateral
 CC sclerosis (ALS), Huntington's disease, retinitis pigmentosa, spinal
 CC muscular atrophy, various forms of cerebellar degeneration, anaemia,
 CC myelodysplastic syndrome, ischemic injury, myocardial infarction,
 CC cerebral ischemia or toxin-induced injury. In addition, T cell
 CC mediated diseases, including AIDS, autoimmune diseases such as
 CC rheumatoid arthritis, and type I diabetes, septic shock, cerebral
 CC malaria, graft rejection, cytotoxicity, cachexia, and inflammation
 CC can be treated by altering the expression or activity of the
 CC polypeptides. The products can also be used for detection,
 CC diagnosis and screening assays.
 SQ Sequence 3964 BP; 981 A; 939 C; 995 G; 1042 T;
 Query Match 7.5%; Score 31; DB 1; Length 3964;
 Best Local Similarity 62.0%; Pred. No. 5.8;
 Matches 49; Conservative 0; Mismatches 30; Indels 0; Gaps 0;
 QY 233 GTCACCTCCACCATCTCAACACCATGTTGTCTGCAACATGACAGCCATTG 292
 Db 3684 GTTACATATAGCTTTGTATATATCCCGCCAAATAGCATGTGCACAGACAGCCATAG 3743
 QY 293 AAGCCTGTGCTCTTTGG 311
 Db 3744 TATATGTGTCACTCGTGG 3762
 RESULT 7
 ID T02964 standard; cDNA; 580 BP.
 AC T02964;
 DT 01-MAR-1996 (first entry)
 DE Cell-cycle regulatory protein p15 cDNA.
 KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
 KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
 KW ss; ds.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT CDS 91..483
 FT /*tag= a
 FN WO9528483-A1.

PD 26-OCT-1995.
 PF 14-APR-1995; US-04636.
 PR 14-APR-1994; US-227371.
 PR 25-MAY-1994; US-248812.
 PR 14-SEP-1994; US-306511.
 PR 29-NOV-1994; US-346147.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PI Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
 DR WPI: 95-373798/48.
 DR P-PSDB; R85118.
 PT New cell cycle regulating proteins bind to cyclin dependent kinase -
 PT and related nucleic acids, antibodies etc., used in diagnosis and
 PT therapy of abnormal cell proliferation, degeneration etc.
 PS Claim 43; Page 81-82; 109pp; English.
 CC cDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
 CC protein p15 (R85118) was isolated from an embryonal carcinoma
 CC library using a probe based on a mouse CCR p13.5 cDNA (T02965).
 CC The isolated cDNA can be used to detect mutations in CCR
 CC genes that lead to cell proliferation; to breed transgenic
 CC animals to study cellular disorders involving CCR allele
 CC mutation/misexpression; and to correct CCR-deficient cells
 CC (gene therapy).
 SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;
 Query Match 7.3%; Score 30.4; DB 1; Length 580;
 Best Local Similarity 57.3%; Pred. No. 4.1;
 Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
 QY 35 GCAGTGGCCACTATGGGTCTGGCTCCCTCTGCTCTTGCACCTCTTGGCAGC 94
 Db 128 GGGTGGCCAGGCCGCGCTACTCTCTCCGCCACACATGCTCTGTCCCGGTCTGTGGC 69
 QY 95 TCACATGAACAGGCCGGGTATGACTTTTGCACACTG 130
 Db 68 AGAATGCTCTTCGCCGCGGTGAGATTGCTACAG 33
 RESULT 8
 ID V50430 standard; cDNA; 1301 BP.
 AC V50430;
 DT 07-DEC-1998 (first entry)
 DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
 KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
 KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT CDS 157..1215
 FT /*tag= a
 FN WO933819-A1.
 PD 06-AUG-1998.
 PF 30-JAN-1998; U01724.
 PR 30-JAN-1997; US-036986.
 PA (UANY) UNIV NEW YORK STATE.
 PI Philipson L, Tomko RP;
 DR WPI: 88-437397/37.
 DR P-PSDB; W69698.
 PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
 PT for preventing and treating viral infection and rendering cells
 PT susceptible to transformation by adenoviral vectors in gene therapy
 PS Claim 2; Page 68-70; 88pp; English.
 CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
 CC serves as a cellular receptor for adenoviruses of the serotypes 2
 CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
 CC was obtained by screening a lambda phage expression cDNA library
 CC with antiserum containing antibodies specific for mouse CVB
 CC binding protein p16, and was used to identify human HCAR cDNA
 CC (see V50429). The invention also provides host cells transformed
 CC with DNA molecules encoding HCAR or MCAR and methods of producing
 CC the recombinant proteins or their derivatives. These proteins,
 CC their extracellular domains, as well as oligopeptides (see
 CC W69699-708) which bind virus, are also provided. Isolated HCAR or


```
CC MCAR proteins or their fragments or variants are used to prevent
CC or treat virus infections. Expressing the DNA in cells which lack
CC these viral receptors renders the cells susceptible to
CC transformation by adenoviral vectors carrying genes for gene
CC therapy.
SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;

Query Match 7.3%; Score 30.4; DB 1; Length 1301;
Best Local Similarity 61.2%; Pred. No. 5.7; Mismatches 0; Gaps 0;
Matches 49; Conservative 0;

QY 29 AGGAGGCGAGTGGCCACTATGGGCTGGGCTGCGCCCTTGCTCTTGACCCCTCTT 88
DB 868 ACGATCGGGGGCGCTCATGAGGCGCTGCTGGCCCTTGCTCATCGGGGCATCCTC 927
QY 89 GCGAGCTCAGTGGACAGG 108
DB 928 TTCGTGTCACAGGAACG 947

RESULT 9
V63094
ID V63094 standard; cDNA; 3881 BP.
AC V63094; 1999 (first entry)
DE Human TR6 cDNA.
KW TR6; tumour necrosis factor related receptor; human; treatment; stroke;
KW inflammation; arthritis; septicaemia; autoimmune disease; restenosis;
KW transplant rejection; infection; ischaemia; brain injury; bone disease;
KW acute respiratory disease syndrome; acquired autoimmune disease syndrome;
KW AIDS; cancer; atherosclerosis; Alzheimers disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS
FT 94..1329
FT /tag= a
FT /product= "TR6"
FT
FT EP-870827-A2.
FT 14-OCT-1998.
FT 23-DEC-1997; 310562.
FT 22-AUG-1997; US-916625.
FT 14-MAR-1997; US-041230.
FT 09-MAY-1997; US-853684.
FT (SMIK ) SMITHKLINE BEECHAM CORP.
FT Deen KC, Young PR;
FT WPI; 98-523156/45.
FT P-PSDB; W76827.
FT DNA encoding tumour necrosis factor receptor TR6 - and corresponding
FT polypeptide, antibody, agonist, antagonist, etc
FT Claim 4; Page 25-27; 34pp; English.
FT This sequence encodes a novel human tumour necrosis factor related
FT receptor, TR6. TR6 polypeptides and polynucleotides can be used in the
FT treatment of chronic and acute inflammation, arthritis, septicaemia,
FT autoimmune diseases (e.g. inflammatory bowel disease, psoriasis),
FT transplant rejection, graft vs. host disease, infection, stroke,
FT ischaemia, acute respiratory disease syndrome, restenosis, brain injury,
FT (acquired autoimmune disease syndrome) AIDS, bone diseases, cancer (e.g.
FT lympho-proliferative disorders), atherosclerosis and Alzheimers disease.
FT Sequence 3881 BP; 937 A; 922 C; 983 G; 1028 T;
SQ

Query Match 7.3%; Score 30.2; DB 1; Length 3881;
Best Local Similarity 59.5%; Pred. No. 10;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 233 GTCACCTCCACCATGATCTCAACACCATGTTGCTGCAACATGACAGCCATTG 292
DB 3651 GTCACATTTGGCTTTGATATATATCCAGGCCAATTCGANGACAGACGACCATG 3710
QY 293 AAGCCTGTGCTCTTCTTGG 311
DB 3711 TATAGTGTGTCACCTCTGG 3729
```

```
RESULT 10
Q74052
ID Q74052 standard; DNA; 7824 BP.
AC Q74052;
DE 29-JAN-1996 (first entry)
DE Human Interleukin-1-beta.
KW Interleukin 1 beta; primer; mRNA; specificity; pharmaceutical; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_difference 1331
FT /tag= a
FT /note= "N is undefined"
FT
FT J07123984-A.
FT 16-MAY-1995.
FT 05-NOV-1993; 275852.
FT 03-NOV-1993; JP-275852.
FT PA (HITB ) HITACHI CHEM CO LTD.
FT DR WPI; 95-211627/28.
FT A primer for the detection and the determ. of a specific messenger
FT RNA - can detect and determine specific mRNA(s) with high
FT reliability
FT Example 18; Page 17-20; 35pp; Japanese.
FT Q74052 is interleukin-1-beta cDNA and Q74019-21 are primers used for
FT the amplification of this cDNA. They are used specifically for the
FT detection and isolation of this sequence. The primers have the
FT advantage of high sensitivity and reliability and are useful in the
FT pharmaceutical industry.
FT Sequence 7824 BP; 2099 A; 1905 C; 1624 G; 2195 T;
SQ

Query Match 7.2%; Score 30; DB 1; Length 7824;
Best Local Similarity 55.9%; Pred. No. 15;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 119 ACTTTCGAAGTGAAGAGCTGTTTCTGTGACAAATTCCTCTATGAGTCCAGCTTC 178
DB 5839 ACTCTAGCTGGAGTGAAGTGAATGATGGAATGAAGCCCTCTCAGCCCTCTGCTAC 5898
QY 179 CTGGAATTCCTTGAAAGGCTCGCTCTCTCTCCATCTCCCT 220
DB 5899 CACTTATCCAGACACACCACTTCTCCCGCCCATCCCT 5940

RESULT 11
T23881/C
ID T23881 standard; cDNA to mRNA; 287 BP.
AC T23881;
DE 16-AUG-1996 (first entry)
DE Human gene signature HUMGS05816.
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW human; cloning; mapping; non-biased library; diagnosis; detection;
KW cell typing; abnormal cell function; ss.
OS Homo sapiens.
PN W09514772-A1.
PD 01-JUN-1995.
PF 11-NOV-1994; J01916.
PR 12-NOV-1993; JP-355504.
PR (MATS/) MATSUBARA K.
PA (OKUB/) OKUBO K.
PI Matsubara K, Okubo K;
DR WPI; 95-206931/27.
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT for diagnosis of abnormal cell function, by preparing cDNA that
PT reflects relative abundance of corresp. mRNA in specific human
PT tissues
PT Claim 1; Page 1477; 2245pp; Japanese.
PS A single-stranded DNA (or its complementary strand or the corresp.
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC given in T19001-T26837 and which is able to hybridise to part of
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC sequences were obtained from 3'-directed cDNA libraries prepared
CC from various human tissues; synthesis of cDNA was initiated from the-
```

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 CC Sequence 287 BP; 63 A; 55 C; 73 G; 81 T;
 SQ

Query Match 7.2%; Score 30; DB 1; Length 287;
 Best Local Similarity 53.4%; Pred. No. 4.1;
 Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 274 CAACATGATGACGACGATGAGCTGTGCTTCTGCGCGGCTTTGGCGGGGAT 333
 DB 132 CAACATGATGACGACGATGAGCTGTGCTTCTGCGCGGCTTTGGCGGGGAT 73
 QY 334 GCAGGAGGAGGCGCGGACCTGTCTTTTCAGCAGGCGGCGGACCTCTGAGTGCGAAT 391
 DB 72 CCATGAGCTAGGCGGACCTGTCTTTTCAGCAGGCGGCGGACCTCTGAGTGCGAAT 15

RESULT 12
 X22303/C
 ID X22303 standard; DNA; 14690 BP.
 AC X22303;
 DT 22-JUN-1999 (first entry)
 DE Human IL-1ra BAC contiguous DNA sequence 96.
 KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
 KW interleukin-1 receptor; IL-18; regulation; asthma; rheumatoid arthritis;
 KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
 KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
 OS Homo sapiens.
 PN W0906426-A1.
 PF 11-FEB-1999.
 PR 03-AUG-1998; U16102.
 PR 02-JUL-1998; US-091650.
 PR 04-AUG-1997; US-054646.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Pan Y;
 DR WPI; 99-153692/13.
 DT New isolated nucleic acid encoding the new human cytokine Tango-77 -
 PT used to inhibit inflammation and to screen for specific modulators
 PS Example 5; Figure 4; 226pp; English.
 CC X02956-X03048 and X2301-X2304 are overlapping BAC genomic sequences
 CC containing alternatively spliced forms of human IL-1ra. Such fragments
 CC are used in the method of the invention which describes the isolation of
 CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
 CC member of the cytokine superfamily that is expected to inhibit
 CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
 CC also bind to a new receptor so could regulate other cellular processes
 CC associated with acute or chronic inflammation, e.g. asthma, chronic
 CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
 CC bowel disease. It may also induce or suppress interleukins, cytokines and
 CC growth factors. Modulators of this protein are used to treat or prevent
 CC conditions associated with abnormal levels of inflammation, or activity
 CC of IL-1 or its receptor complex.
 CC Sequence 14690 BP; 3810 A; 3245 C; 3299 G; 4327 T;
 SQ

Query Match 7.2%; Score 30; DB 1; Length 14690;
 Best Local Similarity 55.9%; Pred. No. 20;
 Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 119 ACTTGCACTGACGAGCTCTTTCTGACAAATCTCTTCATGATGCTCAGCTTC 178
 DB 4343 ACTTACAGCTGGAGGTGAGTGTGCTATGGAATGAAGCCCTCTCAGCCTCTGCTAC 4284
 QY 179 CTGGAATGCTTGAAAGCTCTGCCTCTCTCCATCTCCCT 220
 DE EST clone AR34.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

DB 4283 CACTTATCCAGAACACCACCTTCTCCCGCCCATCCCT 4242
 RESULT 13
 T18696
 ID T18696 standard; cDNA; 3510 BP.
 AC T18696;
 DT 05-JUL-1996 (first entry)
 DE RAP-1 radiation protecting checkpoint protein cDNA.
 KW RAP-1; radiation protecting checkpoint protein; apoptosis;
 KW cell death; cancer; diagnosis; therapy; radiotherapy; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 176..2122
 FT /*tag= a
 FT misc_rna 31077..3510
 FT /*tag= b
 FT /*label= cDNA3
 FT misc_difference 3648
 FT /*tag= b
 FT /*note= "base n at position 3648 is not identified
 FT in the specification"
 PN W09611562-A2.
 PD 25-APR-1996.
 PR 11-OCT-1995; U12445.
 PR 11-OCT-1994; IL-111238.
 PA (SHOS/) SHOSHAN H Z.
 PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
 PI Canaanani D;
 DR WPI; 96-221643/22.
 DR P-PSDB; R94906.
 PT New gene encoding a radiation protecting checkpoint protein - useful
 PT for diagnosis and treatment of cancer and other diseases involving
 PT abnormal apoptosis
 PS Claim 2; Fig 3; 29pp; English.
 CC A cDNA clone (T18696) codes for a novel radiation-protecting human
 CC checkpoint protein (R94906), designated RAP-1, involved in
 CC regulation of cell cycle progression and/or apoptosis. It was
 CC isolated from a human cDNA library established in an immortalised
 CC xeroderma pigmentosa cell line, GM2096 (XP1MI). A cDNA (cDNA3) contg.
 CC a 462 bp insert complemented UV sensitivity and was used to screen a
 CC lambda-g10-X562 human cDNA library. The insert represented part of
 CC the 3510 bp RAP-1 mRNA. The human checkpoint gene can be used as a
 CC marker for early detection and prevention of tumour progression or
 CC can be used to design new drugs that enhance radiotherapy or
 CC chemotherapy of diseases involving abnormal apoptosis.
 CC Sequence 3510 BP; 974 A; 795 C; 848 G; 892 T;
 SQ

Query Match 7.2%; Score 29.8; DB 1; Length 3510;
 Best Local Similarity 51.9%; Pred. No. 13;
 Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 11 CAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGGCTGCCCTTGT 70
 DB 1771 CACCGTCCCTCCATGGAGAGCGAGAGAAATCAATCTCTATCTCTCTCTTGA 1830
 QY 71 CTCCTCTGACCTCTCTGGAGCTCAATGGAACAGCGCGGCTATGACTTTGCACTG 130
 DB 1831 TACCTCTCTGGACTTCTCCAAAGAAACAGAAAAAGGAGGATCTAGTTGGCAGCTT 1890
 QY 131 AAGCTGAAG 139
 DB 1891 AAACGGAGG 1899

RESULT 14
 V86365/c
 ID V86365 standard; cDNA; 291 BP.
 AC V86365;
 DT 27-APR-1999 (first entry)
 DE EST clone AR34.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2000, 16:07:56 ; Search time 61.26 Seconds
(without alignments)
811.048 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419

Perfect score: 415

Sequence: 1 ACGGACTTCAGTGTCTCC.....AAATTCGGTATGCTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 214294 seqs, 59861208 residues

Database : Issued_Patents_NA:*

Word size : 0

Number of hits that pass the threshold : 428588

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/PTUS9_COMB.seq.*
- 7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.2	10.2	7218	1	US-08-232-463-14
2	30.4	7.3	580	4	US-08-627-610-7
3	30.4	7.3	580	6	PCT-US95-04636-7
4	29.8	7.2	3509	3	US-08-817-436A-1
5	29.2	7.0	1560	1	US-07-669-171-1
6	29.2	7.0	5962	1	US-08-188-582-10
7	29.2	7.0	5962	1	US-08-646-715-10
8	29	7.0	3060	1	US-08-098-141-1
9	29	7.0	1157	1	US-07-709-949-1
10	29	7.0	2259	7	5185254-3
11	28.6	6.9	2662	5	US-08-750-357-8
12	28.4	6.8	6306	2	US-08-466-390-3
13	28.4	6.8	6306	2	US-08-470-950-3
14	28.4	6.8	6306	2	US-08-467-781-3
15	28.4	6.8	6306	2	US-08-195-487-3
16	28.4	6.8	6306	3	US-08-483-924-3
17	28.4	6.8	2262	4	US-08-674-887A-5
18	28.4	6.8	6306	6	PCT-US93-06160-3
19	27.8	6.7	185	3	US-08-332-766A-31
20	27.8	6.7	2782	3	US-08-937-466-1
21	27.8	6.7	2782	4	US-09-172-528-1
22	27.6	6.7	1569	1	US-07-669-171-3
23	27.6	6.7	1853	2	US-08-553-110-2
24	27.6	6.7	5102	2	US-08-494-168-1
25	27.6	6.7	3088	2	US-08-418-444A-1
26	27.6	6.7	3282	3	US-08-678-039A-41
27	27.4	6.6	1364	2	US-08-306-691B-50
28	27.4	6.6	4425	3	US-08-749-169A-1
29	27.4	6.6	2865	3	US-08-749-169A-2
30	27.4	6.6	4425	4	US-09-130-032A-1
31	27.4	6.6	2865	4	US-09-130-032A-2
32	27.4	6.6	3037	4	US-08-938-365-1

c 33	27.4	6.6	1364	6	PCT-US93-06251-65	Sequence 65, Appli
c 34	27.2	6.6	2043	1	US-07-914-281-1	Sequence 1, Appli
c 35	27.2	6.6	2043	1	US-08-393-246-1	Sequence 1, Appli
c 36	27.2	6.6	2043	1	US-08-273-411-4	Sequence 4, Appli
c 37	27.2	6.6	2121	1	US-08-331-394-3	Sequence 3, Appli
c 38	27.2	6.6	2121	2	US-08-250-858-3	Sequence 3, Appli
c 39	27.2	6.6	2121	2	US-08-448-915-3	Sequence 3, Appli
c 40	27.2	6.6	2043	2	US-08-525-058A-1	Sequence 1, Appli
c 41	27.2	6.6	2122	2	US-08-485-449-1	Sequence 1, Appli
c 42	27.2	6.6	2121	3	US-08-744-139-3	Sequence 3, Appli
c 43	27.2	6.6	2043	4	US-08-696-731-1	Sequence 1, Appli
c 44	27.2	6.6	2043	6	PCT-US91-00899-1	Sequence 1, Appli
c 45	27.2	6.6	2121	6	PCT-US95-06639-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 10.2%; Score 42.2; DB 1; Length 7218;
Best Local Similarity 2.8%; Pred. No. 0.0008;
Matches 11; Conservative 214; Mismatches 162; Indels 0; Gaps 0;
QY 4 GGGACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCATGATGGGTCTGGGCTGCC 63


```
; FILING DATE: 19930723
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kaplan Esq., Warren A.
; REGISTRATION NUMBER: 34,199
; REFERENCE/DOCKET NUMBER: 092662-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617:227-5020
; TELEFAX: 617:227-7566
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-098-141-1

Query Match          7.0%; Score 29; DB 1; Length 3060;
Best Local Similarity 50.4%; Pred. No. 7.7; Mismatches 0; Indels 0; Gaps 0;
Matches 71; Conservative 0;

QY 202 CTTCTCTTCATCTCCCTTCAGGAGCAGCGTCACCTCCACCATGCAAGATCTCAACA 261
Db 1540 CCAGTCTTTCACATGCGCTTCAACATGCGCGTGACCTGTACTGTGGCAGCCACAGGCCA 1599

QY 262 CCAATGTTGTGCAACATGACAGCCATTGAAGCCTGTGCTCTTGTGGCCCGGCTTT 321
Db 1600 CTACACCTCTCTTCTCCCAACACACTGCTGCAGCCTGTATCTCTCGTGCCCAACATCAC 1659

QY 322 TGGCGGGGGATGACGAGGC 342
Db 1660 CTGGTCAGAGATCCAGGTGCC 1680

RESULT 9
US-07-709-949-1
; Sequence 1, Application US/07709949
; Patent No. 5472858
; GENERAL INFORMATION:
; APPLICANT: Attie, Alan D
; APPLICANT: Gretch, Daniel G
; APPLICANT: Sturley, Stephen L
; APPLICANT: Beckage, Nancy E
; TITLE OF INVENTION: Production of Recombinant Proteins in
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/709,949
; FILING DATE: 19910604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 9629691801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62..1015
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 62..119
; PUBLICATION INFORMATION:
; AUTHORS: McLean, J W
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 259
; PAGES: 6498-6504
; DATE: 1984
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1157
; US-07-709-949-1

Query Match          7.0%; Score 29; DB 1; Length 1157;
Best Local Similarity 54.18; Pred. No. 5; Mismatches 0; Indels 0; Gaps 0;
Matches 59; Conservative 0;

QY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCTGCCCC 65
Db 626 GAGCGCGCGCTCAGCGCCATCCGCGAGCGCTGGGCGCCCTGTGTGAACAGGCGCGCTG 685

QY 66 TTGTCTCTCTTTGACCTCTTGTGGAGCTCACATGGAACAGAGCGCGGG 114
Db 686 CGGCGCGCCACTGTGGGCTCTCCCTGGCGCGCCAGCGCTACAGAGAGCGGG 734

RESULT 10
5185254-3
; Patent No. 5185254
; APPLICANT: LINNENBACH, ALBAN
; TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/291,583
; FILING DATE: 29-DEC-1988
; SEQ ID NO:3:
; LENGTH: 2259
; 5185254-3

Query Match          7.0%; Score 29; DB 7; Length 2259;
Best Local Similarity 52.0%; Pred. No. 6.8; Mismatches 0; Indels 0; Gaps 0;
Matches 65; Conservative 0;

QY 17 CTCCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCTGTGCTCTCTC 76
Db 1120 qcctcaccgcggcctcatcgctcgctggtggtcggtggtggtcgctcgccgcgc 1179

QY 77 TTGACCTCTTGGCAGCTCACATGGACAGCGCGGTATGACTTTGCNACTGAAGCTG 136
Db 1180 atggccgtcctggtgatcaccacccgggaaagtgcgggaaagtacagaagtggagatc 1239

QY 137 AAGGA 141
Db 1240 aagga 1244

RESULT 11
```

US-08-750-357-8/c
; Sequence 8, Application US/08750357
; Patent No. 6008437
; GENERAL INFORMATION:
; APPLICANT: KREBBERS, Enno
; APPLICANT: WILLIAMS, Mark
; APPLICANT: LEEMANS, Jan
; TITLE OF INVENTION: USE OF ANTHOCYANIN GENES TO MAINTAIN
; TITLE OF INVENTION: MALE STERILE PLANTS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/750.357
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 018030-100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2662 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-750-357-8

Query Match 6.9%; Score 28.6; DB 5; Length 2662;
Best Local Similarity 61.3%; Pred. No. 9.7;
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY 203 CTCCTCCTCATCTCCCTTCAGGACGAGCGTCCACCTCCACCATGCAAGATCTCAACAC 262
Db 1212 CTCCTTCTCCACCGCGTTCCAGGTCGCGCGACGCGCACACCAAGGACCTCCACAA 1153
QY 263 CATGTTGTCGCAAC 277
Db 1152 AAACGGCAGCAAC 1138

RESULT 12
US-08-466-390-3/c
; Sequence 3, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466.390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992
US-08-466-390-3

Query Match 6.8%; Score 28.4; DB 1; Length 6306;
Best Local Similarity 46.1%; Pred. No. 16;
Matches 95; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 188 CTTGAAGCTCTGCTCCTCCTCCATCTCCCTTCAGGACGAGCGTCCACCTCCACCAT 247
Db 4728 CTGGAAGCGCTGGGCGCTCCTGCTGCTCCTCCCTCCTGAGCGTGGACGCTTCAGCTT 4669
QY 248 GCAAGATCTCAACACCATCTTGTCTGCAACACATGACGCCATTGAAGCCTGTGCTTTC 307
Db 4668 CTGCTGCTGCACCTTCTGCTGGCTTGGTCAGAGTCAGCCAGTTTCTTACTCAGTTCTTCCAC 4609
QY 308 TTGGCCCGGGGCTTTTGGCGGGGATGACGAGGACGAGCGGCGGCGGCGGCGGCGGCGG 367
Db 4608 CTGGCAGTCAGTTTCTGCTCCTCTCTCTGGAACCGCTGCTCTCTCTCCAGGACCTTGAC 4549
QY 368 GCCCCACCTCCTCCTGAGTGGCAATAA 393
Db 4548 CTGGCACCTCCTATCTTGGCAGTCA 4523

RESULT 13
US-08-470-950-3/c
; Sequence 3, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2000, 08:56:37 ; Search time 538.4 Seconds
(without alignments)
2910.291 Million cell updates/sec

Title: US-09-092-296-7_COPY_5_419
Perfect score: 415
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....AAAATCGGTATCGTGAATT 415

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	308.4	74.3	404	61	AI857998
c 2	153	36.9	552	82	AI857998 w169b01.x
c 3	152.8	36.8	176	21	AQ18761 HS_5511.B
c 4	106	25.5	328	42	T94049 ye33f07.r1
c 5	46.8	11.3	190	41	AI136523 UI-R-C2p-
c 6	40.8	9.8	201	47	AI010074 EST204525
c 7	37.4	9.0	1101	79	AI535335 UI-R-C3-s
					AL076960 Drosophil

```
c 8 36.4 8.8 331 31 AA282040
c 9 36.4 8.8 494 34 AA534569
c 10 36.4 8.8 424 38 AA741185
c 11 36.4 8.8 402 40 AA936790
c 12 36.4 8.8 529 41 A1018775
c 13 35.6 8.6 354 70 AW143645
c 14 35.2 8.5 537 33 AA417974
c 15 35.2 8.5 460 40 AA909652
c 16 35 348 21 F06958
c 17 34.8 8.4 451 61 A1829206
c 18 34.8 8.4 939 79 CNS00CNG
c 19 34.4 8.3 378 30 AA243870
c 20 34.4 8.3 238 32 AA376266
c 21 34.4 8.3 329 38 AA769413
c 22 34.4 8.3 329 38 AA769413
c 23 34.4 8.3 412 38 AA769782
c 24 34.4 8.3 458 49 A1632159
c 25 34.4 8.3 470 60 A1803529
c 26 34.2 8.2 509 88 AQ838514
c 27 33.6 8.1 429 39 AA862672
c 28 33.4 8.0 258 20 D34884
c 29 33.4 8.0 545 23 T41524
c 30 33.2 8.0 292 40 AA91822
c 31 33.2 8.0 597 63 AW013553
c 32 33 8.0 261 29 AA183985
c 33 32.8 7.9 405 21 R02824
c 34 32.8 7.9 360 35 C43492
c 35 32.8 7.9 240 36 C70784
c 36 32.6 7.9 239 39 AA830716
c 37 32.6 7.9 522 82 AQ718755
c 38 32.4 7.8 490 51 A1757641
c 39 32.4 7.8 537 64 AW078074
c 40 32.4 7.8 523 69 AW128453
c 41 32.2 7.8 339 21 F08745
c 42 32.2 7.8 742 27 Z78408
c 43 32.2 7.8 477 45 A1351147
c 44 32.2 7.8 250 62 A1926123
c 45 32.2 7.8 875 82 AQ740265
```

ALIGNMENTS

```
RESULT 1
LOCUS wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
DEFINITION mRNA sequence.
ACCESSION AI857998
VERSION AI857998.1 GI:5511614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.lnl.gov/bbrp/image/Image.html
```

```
Seq primer: -40UP from Gibco
High quality sequence stop: 395.
FEATURES
source
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2408041"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="organ: lung; Vector: p7T3D-Pac (Pharmacia) with a
modified polylinker: 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 74.3%; Score 308.4; DB 61; Length 404;
Best Local Similarity 99.4%; Pred. No. 7.2e-79;
Matches 309; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 105 CAGGCGCGGGTATGACTTTGCAACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 164
Db 311 CAGGCGCGGGTATGACTTTGCAACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 252
QY 165 ATGAGTCAGCTTCCTGGAATTCCTGAAAGCTCGCTCTCTCCATCTCCCTTCAG 224
Db 251 ATGAGTCAGCTTCCTGGAATTCCTGAAAGCTTCCTCTCCATCTCCCTTCAG 192
QY 225 GGACGAGGTCACCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACACATGAC 284
Db 191 GGACGAGGTCACCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACACATGAC 132
QY 285 AGCCATTGAGCCTGTCTCTTCTTGGCCCGGGCTTTTGGCCCGGGATCGAGGCGAG 344
Db 131 AGCCATTGAGCCTGTCTCTTCTTGGCCCGGGCTTTTGGCCCGGGATCGAGGCGAG 72
QY 345 GCCCGAGCCTGTCTTTCAGCAGGCCCCACCTCTCTGAGTGGCAATAAAATAAATTCGG 404
Db 71 GCCCGAGCCTGTCTTTCAGCAGGCCCCACCTCTCTGAGGCGCAATAAAATAAATTCGG 12
QY 405 TATGCTGAAT 415
Db 11 TATGCTGAAT 1
RESULT 2
LOCUS AQ718761/c
DEFINITION HS_5511_B2_F09_11 Human Male BAC Library Homo sapiens
genomic clone Plate-1087 Col-18 Row-L, genomic survey sequence.
ACCESSION AQ718761
VERSION AQ718761.1 GI:5468077
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
```


451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics

Seq primer: M13 Forward.

FEATURES

Location/Qualifiers
 1. 328
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C2p-qq-e-02-0-UI"
 /clone_lib="UI-R-C2p"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-805, 1996)"

BASE COUNT 62 a 77 c 98 g 91 t

Query Match 25.5%; Score 106; DB 42; Length 328;

Best Local Similarity 66.8%; Pred. No. 1.8e-20;

Matches 151; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 90 GCACCTCACATGGACAGCGCGGGTATGACCTTTCGACACTGACGCTGAAGGAGTCTTTC 149

Db 327 GCAGCTCAGTCGAGGGCCACAAAGGTGACCTTTCGAAGTAAACACTGACCGAGGACATCTC 268

QY 150 TGACAAATTCCTCTATGAGTCCAGCTTCTCTGGAATTCCTTGAAGGCTCTCGCTCTCTCC 209

Db 267 AGGCCAAGACCTCCCAAGACTCCGGCTTTCTGGACATGCTCCAAAGATCTCCCTCTCC 208

QY 210 TCCATCTCCCTCAGGACGACGCTACCTCCACCATGCAAGATCTCAACCATCTTG 269

Db 207 TCCACCTCTCAGCGGGACCAATGTCACCTTCATCATAAAGGCCACACACACCTTA 148

QY 270 TCTGCACACATGACAGCCATTAAGCGCTGTGTCTCTTTGGCCCG 315

Db 147 CCTGCAGAGCTGAGAGCAGCTGAAGCGCTGCTGGGGCTGGAGCGG 102

RESULT 5

AI010074/c

LOCUS

DEFINITION

UIRBT38 3' end, mRNA sequence.

ACCESSION

AI010074

190 bp mRNA

EST

15-JUN-1998

AI010074.1 GI:3223906

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Rattus sp.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 190)
 Lee,N.H., Glodex,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Jan 19, 1998 this sequence version replaced gi:2150555.
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES

Location/Qualifiers

1. 190

/organism="Rattus sp."

/db_xref="taxon:10118"

/clone_lib="Normalized rat lung, Bento Soares"

/note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI;

Site_2: NotI"

BASE COUNT 39 a 51 c 59 g 41 t

ORIGIN

Query Match 11.3%; Score 46.8; DB 41; Length 190;

Best Local Similarity 65.1%; Pred. No. 0.0019;

Matches 69; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 210 TCCATCTCCCTCAGGACGACGCTCACCTCCACCATGCAAGATCTCAACCATCTTG 269

Db 190 TCCACCTCTCACCAGGACCAATGTCACCTTCATCATAAAGGCCACACACACCTTA 131

QY 270 TCTGCACACATGACGACCATTAAGCGCTGTGTCTCTTTGGCCCG 315

Db 130 CCTGCAGAGCTGAGAGCTGAGCGCTGCTGGGGCTGGAGCGG 85

RESULT 6

AI535335/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AI535335 201 bp mRNA EST 18-MAR-1999
 UI-R-C3-sw-e-12-0-UI.s1 UI-R-C3 Rattus norvegicus cDNA clone
 UI-R-C3-sw-e-12-0-UI 3', mRNA sequence.

AI535335
 AI535335.1 GI:4449470
 EST.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 201)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)

On May 18, 1998 this sequence version replaced gi:3136861.
 9704477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: mscoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1402231.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 891 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 242.

FEATURES
source

1. .331
/organism="Homo sapiens"
/db_xref="GDB:5854344"
/db_xref="taxon:9606"
/clone="IMAGE:704620"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-).
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGGGAGCGCGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN

97 a 61 c 51 g 122 t

Query Match 8.8%; Score 36.4; DB 31; Length 331;
Best Local Similarity 56.8%; Pred. No. 2.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GCTGAAGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTTCCTGGAATGCTTGA 192
|||||
Db 318 GATGATGACTACTGTCTATATAATCACTAAATCCAGCTACCGAGAACTGCCTGG 259
|||
QY 193 AAAGCTCTGCCCTCCCTCCATCCTCCCTCAGGACCGAGCTCACCCTCCACCATGCA 250
|||
Db 258 AACTGTGCCATGCATTTTTTTCTTTAAAGACCAAGTGTGATAGTAGGCCATGCA 201

RESULT 9
AA534569/c

LOCUS n177c11.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:925940 3',
DEFINITION mRNA sequence.
ACCESSION AA534569
VERSION AA534569.1 GI:2278822
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 494)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index

JOURNAL

Unpublished (1997)

On May 9, 1995 this sequence version replaced gi:802984.

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 370.

FEATURES
source

1. .494
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:925940"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"

/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library went through one round of
normalization."

BASE COUNT 139 a 108 c 94 g 153 t

ORIGIN

Query Match 8.8%; Score 36.4; DB 34; Length 494;
Best Local Similarity 56.8%; Pred. No. 2.6;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GCTGAAGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTTCCTGGAATGCTTGA 192
|||||
Db 296 GATGATGACTACTGTCTATATAATCACTAAATCCAGCTACCGAGAACTGCCTGG 237
|||
QY 193 AAAGCTCTGCCCTCCCTCCATCCTCCCTCAGGACCGAGCTCACCCTCCACCATGCA 250
|||
Db 236 AACTGTGCCATGCATTTTTTTCTTTAAAGACCAAGTGTGATAGTAGGCCATGCA 179

RESULT 10

AA741185/c
LOCUS ob30a02.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1325162 3',
DEFINITION mRNA sequence.

ACCESSION AA741185

VERSION AA741185.1 GI:2779777

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 424)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished (1997)

On Jan 5, 1998 this sequence version replaced gi:2581148.

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
unknown library type
Insert Length: 742 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 387.

FEATURES
source

1. .424
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1325162"

/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', ACTGAGAGATTCGGCGGCAATATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 121 a 90 c 77 g 136 t
ORIGIN

Query Match 8.8%; Score 36.4; DB 38; Length 424;
Best Local Similarity 56.8%; Pred. No. 2.5;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GCTGAGGAGCTCTTTCTGACAAATCTCTATGAGTCACGCTCTCGGAATGCTTGA 192
|||||
DB 303 GATGATGACTGCTACTGCTATATAAATCCACTAAATCCAGCTACCGAGAACTGCTGG 244
|||||

QY 193 AAGCTCTGCTCTCTCTCCATCTCCTCAGGACGACGCTCCACCATGCA 250
|||||
DB 243 AACTGTGGCCATGCAATTTTTTTCTTTAAAGACCACTGTGATAGTAGGCCATGCA 186
|||||

RESULT 11
LOCUS AA936790 402 bp mRNA EST 19-MAY-1998
DEFINITION chr89a02.s1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1474154 3', mRNA sequence.
ACCESSION AA936790
VERSION AA936790.1 GI:3094824
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045507.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 676 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 386.
Location/Qualifiers
FEATURES
source
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I -

oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 118 a 85 c 74 g 125 t
ORIGIN

Query Match 8.8%; Score 36.4; DB 40; Length 402;
Best Local Similarity 56.8%; Pred. No. 2.5;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GCTGAGGAGCTCTTTCTGACAAATCTCTATGAGTCACGCTCTCGGAATGCTTGA 192
|||||
DB 296 GATGATGACTGCTACTGCTATATAAATCCACTAAATCCAGCTACCGAGAACTGCTGG 237
|||||

QY 193 AAGCTCTGCTCTCTCTCCATCTCCTCAGGACGACGCTCCACCATGCA 250
|||||
DB 236 AACTGTGGCCATGCAATTTTTTTCTTTAAAGACCACTGTGATAGTAGGCCATGCA 179
|||||

RESULT 12
LOCUS AI018775 529 bp mRNA EST 27-AUG-1998
DEFINITION ov32e12.x1 soares_testis_NHT Homo sapiens cDNA clone IMAGE:1639054 3', mRNA sequence.
ACCESSION AI018775
VERSION AI018775.1 GI:3232573
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 14, 1998 this sequence version replaced gi:1797652.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1493 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 451.
Location/Qualifiers
FEATURES
source
1. .529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1639054"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5', TGTTCACCAATCTGAAGTGGGCGCGCCCAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to C05, and was constructed by Bento Soares and M. Fatima Bonaldo."

```

BASE COUNT      146 a   114 c   104 g   165 t
ORIGIN

Query Match      8.8%; Score 36.4; DB 41; Length 529;
Best Local Similarity 56.8%; Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 133 GCTGAAGAGTCTTTCTTGACAAATCTCTATGAGTCCAGCTTCTGGAATGCTTGA 192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 GATGATGACTGACTGATATATTAATCACTAAATCCAGTACCAGGAAGTGCCTGG 240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 193 AAAGCTGCTGCTCTCTCCATCTCCCTTCAGGACCAAGCTCACCTCCACCATGCA 250
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 AACTGTGGCATGCAATTTTTTTTTTTCTTTAAAGACCAAGTGTGATAGTGGCATGCA 182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AW143645/c
LOCUS
DEFINITION
  EST293941 Normalized rat embryo, Bonto Soares Rattus sp. cDNA clone
ACCESSION
  AW143645
VERSION
  AW143645.1 GI:6163544
KEYWORDS
  EST.
SOURCE
  Rattus sp.
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
  1 (bases 1 to 354)
  Lee N.H., Glodek A., Chandra, I., Mason, T.M., Quackenbush, J.,
  Kerlavage, A.R. and Adams, M.D.
  Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat
  Gene Index
  Unpublished (1998)
  On Apr 30, 1999 this sequence version replaced gi:4727610.
  Contact: Lee, NH
  ATCC

The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability, additional sequence and expression
information related to this EST please check the TIGR Rat Gene
Index (http://www.tigr.org/tdb/rat/rat.html). To order a clone
contact the ATCC (http://www.atcc.org/atcc.html).
Seq primer: M13 Reverse.
FEATURES
  source
    1..354
    /organism="Rattus sp."
    /db_xref="taxon:10118"
    /clone="RG1Bx27"
    /dev_stage="embryo 8, 12, 18 dpc"
    /note="Vector: pT73pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      106 a   76 c   82 g   90 t
ORIGIN

Query Match      8.6%; Score 35.6; DB 70; Length 354;
Best Local Similarity 45.9%; Pred. No. 4;
Matches 122; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 118 GACTTTGCACTGAAGCTGAAGAGTCTTTCTGACAAATTCCTCTATGATGCTCAGCTT 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 GATTTTCTTCCTTAACTGCAATCTACTTTCGAGAAAGATCTCAAGTGTCTGTGAGCT 210
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 178 CTGGAATGCTTGAAAGCTCTGCTCCCTCCATCTCCCTTCAGGACAGCGGTCAAC 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 CCAACACCTGATTAAGTCTGAGTCTTTCTTCCAGTTCCACATGATTTTTCAGCATCA 150
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 238 CCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACACATGACAGCATCAAGCC 297

```

```

Db 149 CTGGTCCATGTCAGCATTCATCTTGGGAACAGACGCTTCAGCTCCGGAATGCAAC 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 TGTGTCCTCTTGGCCCGGCTTTTGGCCGGGATGACGAGGAGCGAGCCCGACCTGT 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 TCTTTAATGAATGGTCTTGGCTTTGAGCGCGCGGAAAGCGGATAGTAGTACGCGCT 30
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 CTTTTCAGCAGCGCCCGCCACCTCTCTGA 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 CAGCGCGGCTCTCCACCACTGA 4
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
AA417974
LOCUS
DEFINITION
  zv97c08.s1 Soares_NhMPu_S1 Homo sapiens cDNA clone IMAGE:767726 3'
  similar to WP:F59B2.3 CE00231 N'-ACETYL-GLUCOSAMINE-6-PHOSPHATE
  DEACETYLASE ; mRNA sequence.
ACCESSION
  AA417974
VERSION
  AA417974.1 GI:2079793
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 537)
  Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
  Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,
  Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,
  White, Y., Wylie, T., Waterston, R. and Wilson, R.
  WashU-Merck EST Project 1997
  Unpublished (1997)
  On Sep 12, 1996 this sequence version replaced gi:1393089.
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (infoimage.llnl.gov) for further information.
  Possible reversed clone: similarity on wrong strand
  Possible reversed clone: polyT not found
  Seq primer: -41m3 fwd. ET from Amersham
  High quality sequence stop: 502.
  Location/Qualifiers
    1..537
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /clone="IMAGE:767726"
    /clone_lib="Soares_NhMPu_S1"
    /tissue_type="Pooled human melanocyte, fetal heart, and
    pregnant uterus"
    /lab_host="DH10B"
    /note="Organ: mixed (see below); Vector: pT73D-Pac
    (Pharmacia) with a modified polylinker; Site_1: Not I;
    Site_2: Eco RI; Equal amounts of plasmid DNA from three
    normalized libraries (melanocyte 2NBHM, pregnant uterus
    NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
    were made in vitro. Following HAP purification, this DNA
    was used as tracer in a subtractive hybridization
    reaction. The driver was PCR-amplified cDNAs from pools of
    5,000 clones made from the same 3 libraries. The pools
    consisted of I.M.A.G.E. clones 260232-265223,
    340488-345479, and 484488-489479."
BASE COUNT      90 a   193 c   166 g   88 t
ORIGIN

Query Match      8.5%; Score 35.2; DB 33; Length 537;
Best Local Similarity 47.7%; Pred. No. 6.1;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

```


REFERENCE
AUTHORS

1. (sites)
Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T.,
Kikkawa, E., Taka, K., Tomizawa, M., Okueki, N., Kuwano, Y.,
Watanabe, K., Fukuzumi, Y., Itakura, S., Sugawara, C., Ono, A.,
Yamada, M., Tashiro, H., Ando, A., Ikemura, T., Soeda, E., Kimura, M.,
Bannai, S., and Inoko, H.
Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1,796,998-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

TITLE

2. (bases 1 to 192650)
Shiina, T. and Takishima, N.
Direct Submission
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases, Takashi
Shiina, Tokai University School of Medicine, Department of
Molecular Life Science 2, Honseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail: tshinai@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)

FEATURES

source
1. .192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SK"
/chromosome="6"
/clone="539g"
/map="6p21.3"
BASE COUNT 49862 a 44743 c 45833 g 52212 t

ORIGIN

Query Match 39.5%; Score 92.4; DB 10; Length 192650;
Best Local Similarity 98.9%; Pred. No. 4e-18; Indels 0; Gaps 0;
Matches 93; Conservative 0; Mismatches 1;
QY 59 CAGGCCGGGTATGACTTTCGAACCTGAAGCTGAAGAGCTCTTTTCGACAAATTCCTCT 118
Db 177979 CAGGCCGGGTATGACTTTCGAACCTGAAGCTGAAGAGCTCTTTTCGACAAATTCCTCT 178038
QY 119 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGCT 152
Db 178039 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGGT 178072

RESULT 2

AP000511 200000 bp DNA PRI 28-SEP-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 10/20.
ACCESSION AP000511
VERSION AP000511.1 GI:5926698
KEYWORDS
SOURCE Homo sapiens DNA.

ORGANISM

Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1. (sites)
Shiina, S., Tamiya, G., Oka, A. and Inoko, H.
Homo sapiens 2,429,818bp genomic DNA of 6p21.3 HLA class I region
JOURNAL Published Only in DataBase (1999) in press
REFERENCE 2 (bases 1 to 200000)
AUTHORS Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, J.
TITLE Direct Submission
JOURNAL Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases, Miki
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail: mika@tokyo.jst.go.jp,
URL: http://www.alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)

COMMENT

This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone: +81-463-93-1121. Fax: +81-463-94-8884
The sequence is submitted by Human Genome Sequencing in ALIS

FEATURES

source
1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774.108864
/standard_name="D6S2088"
/note="SHGC-12985; The location is between each flanking
site of PCR primers."
/db_xref="GDB:735268"
113646.113906
/standard_name="D6S1898"
/note="WI-9418; The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
complement(113665..115896)
/note="CDA0VH10; The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"
142030.142159
/standard_name="D6S2203"
/note="SHGC-16870; The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"
complement(142085..142322)
/note="RH1813; The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479..142740)
/standard_name="D6S1851"
/note="SHGC-10808; The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"
complement(164525..164665)
/note="SHGC-3064; The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209..169574)
/standard_name="D6S952"
/note="U75233; The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"

BASE COUNT

52605 a 47531 c 49366 g 50498 t

ORIGIN

Query Match 39.5%; Score 92.4; DB 10; Length 200000;
Best Local Similarity 98.9%; Pred. No. 4e-18;
Matches 93; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 59 CAGGCCGGGTATGACTTTCGAACCTGAAGCTGAAGAGCTCTTTTCGACAAATTCCTCT 118
Db 110306 CAGGCCGGGTATGACTTTCGAACCTGAAGCTGAAGAGCTCTTTTCGACAAATTCCTCT 110365
QY 119 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGCT 152
Db 110366 ATGAGTCCAGCTTCCTGGAATTCGTTGAAAGGT 110399

RESULT 3

AC005937 47323 bp DNA PRI 05-NOV-1998
LOCUS Homo sapiens clone UNGC:370M23.002 from 6p21, complete sequence.
DEFINITION AC005937
ACCESSION AC005937
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eutrophia; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Buticchia; Primates; Hominidae; Hominidae; Homo.
1 (bases 1 to 47323)
Geraghty, D. E. Sequence analysis of the human MHC class I region
TITLE Published (1998)
JOURNAL The Clifton Research Center
REMARK The Clifton Research Center
1100 Fairview, N., P.O. Box 19024
Seattle, WA 98108-1024
2 (bases 1 to 47323)
Geraghty, D. E. and Olson, M. V.
DIRECT SUBMISSION
TITLE Submitted (05-NOV-1998) Human Genome Center, University of
JOURNAL Washington, Box 352145, Seattle, WA 98195, USA
REMARK University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhrc.org)
COMMENT Overlapping Sequences
5': UMGCG:370M23.013 (Genbank Accession: AC005530)
3': UMGCG:Y67C112 (Genbank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All mappings to the reference sequence have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%
DS of two chemistry coverage: 98.9%
Single stranded regions:

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
Mapping. Comparison of the experimentally derived map digest
fragments with sequence-predicted fragments is given below.
Small fragments below a variable cutoff (approximately 400-600bp)
are not mapped and hence do not appear in the table. There are no
significant remaining discrepancies between the experimental and
predicted values. Uniquely ordered fragment groups are separated
by dashed lines.

Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00
20320.67	20855.00	1050.18	1015.00
2171.50	2147.00	7268.78	7196.00
2560.20	2531.00	10085.80	9992.00
4335.42	4269.00	11212.78	11131.00
2698.62	2628.00		
1927.50	1887.00		
3130.46	3090.00		
2166.69	2129.00		
2044.67	2005.00		

Location/Qualifiers
1. 47323

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/clone_lib="Research Genetics BAC Library"
3647..3932
/rpt_family="Alu"
complement(4999..5277)
/rpt_family="Alu"
6285..6572
/rpt_family="Alu"
complement(6972..7050)
/rpt_family="MUT1"
7286..7384
/rpt_family="Alu"
complement(8164..8609)
/rpt_family="Alu"
complement(21287..21895)
/rpt_family="Alu"
22715..22957
/rpt_family="Alu"
25510..25802
/rpt_family="Alu"
27835..28010
/rpt_family="MER20"
31295..31594
/rpt_family="Alu"
33515..33767
/rpt_family="Alu"
34223..34390
/rpt_family="MIR"
37372..37648
/rpt_family="Alu"
38526..38700
/rpt_family="MER3"
39583..40010
/rpt_family="Alu"
40046..40156
/rpt_family="Alu"
43194..43372
/rpt_family="MERS"
43325
/note="clonal variation with 3' overlapping clone"
44149
/note="clonal variation with 3' overlapping clone"
44451
/note="clonal variation with 3' overlapping clone"
44537
/note="clonal variation with 3' overlapping clone"
44814
/note="clonal variation with 3' overlapping clone"
44965
/note="clonal variation with 3' overlapping clone"
45760
/note="clonal variation with 3' overlapping clone"
45900
/note="clonal variation with 3' overlapping clone"
46851
/note="clonal variation with 3' overlapping clone"
46859
/note="clonal variation with 3' overlapping clone"
47032
/note="clonal variation with 3' overlapping clone"
47240..47256
/note="clonal variation with 3' overlapping clone"
insertion of 17bp repeat
BASE COUNT 11556 a 11489 c 12284 g 11994 t
ORIGIN

Query Match 39.5%; Score 92.4; DB 11; Length 47323;
Best Local Similarity 98.9%; Pred. No. 3.7e-18;

```

Matches 93:  Conservative  0:  Mismatches  1:  Indels  0:  Gaps  0:

```

59	QY	CAGGCGCGGTATGACTTTGCACCTGAAGCTCTTTTCTGCAAAATTCCTCT	118
3465	DB	CAGGCGCGGTATGACTTTGCACCTGAAGCTCTTTTCTGCAAAATTCCTCT	3524
119	QY	ATGAGTCGCAGCTTCTGGAATTCGTTGAAAGCT	152
35325	DB	ATGAGTCGCAGCTTCTGGAATTCGTTGAAAGCT	35558

RESULT	4	PRI	28-AUG-1998
AC004148/c	118276 bp	DNA	
LOCUS			
DEFINITION	Homo sapiens chromosome 17, clone HCT524C5, complete sequence.		
AC004148			
ACCESSION			
VERSION	1		
AC004148.1	GI:3482960		

REINFORCED SOURCE ORGANISM
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 118276)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 17, clone HCU524C5
 TITLE

2 (bases 1 to 118276)

Blirg, B., Fasman, K., McKernan, K., Nusbbaum, C., Richardson, P., Lander, E., Algen, N., Baker, J., Baldwin, J., Barna, N., Becker, R., Boutwell, C., Byrne, S., Cantu, C., Castile, A., Certy, D., Cooke, P., Daly, M. J., Depayre, E., Devou, K., Dewar, K., Donelan, L., Dukette, B., Etemadi, S., Ferrel, P., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Gerargery, K., Gilmartin, T., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J. C., Hul, L., Jacotot, L., Linton, L., MacKenzie, J., Marquis, N., McEwan, P., McGurk, A., Meldrim, J., Molla, M., Morris, M., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Ranganathan, S., Riley, R., Roberts, R., Rollins, G., Rossello, R., Roy, A., Shyam, R., Sochoo, S., Stange-Thomann, N., Stillwell, J., Stone, C., Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Ye, W. J., Zemtseva, I., Zhao, J. and Zody, M.

TITLE
JOURNAL
REFERENCE

AUTHORS

Birren, B., Fisman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Bann, J., Boutwell, C., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Corliss, D., Depavate, E., Devon, K., Dewar, K., Donelan, L., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gaugerly, K., Grant, G., Hagos, B., Heatford, A., Herena, L., Horton, L., Lechowzy, J. C., Jacotot, L., Jones, C., Kann, L., Karatas, A., Lechowzy, K., Macdonald, P., Marquis, N., McSwan, P., McGurk, A., McKernan, J., Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naif, R., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Paylin, B., Peterson, K., Riley, R., Roberts, D., Roy, A., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Tichovskiy, N., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D., Ye, Y., Zhao, J. and Zody, M.

TITLE	Direct Submission
JOURNAL	Submitted (28-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	On Aug 28, 1998 this sequence version replaced gi:3402743. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

```

FEATURES
  source
    1. .118276
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="HCII524C3"

```

```

/cytosine.lib--Research Genetics/Cat Tech CTRB-HSP-C (plate)
/mag="814"
/chromosome="17"
complement(295..592)
/rpt.family="AlusX"
complement(1063..1204)
/rpt.family="L1MB8"
1229..1276
/rpt.family="(CA)n"
complement(1273..1468)
/rpt.family="L1MB8"
complement(1512..1852)
/rpt.family="AlusX"
complement(1975..2142)
/rpt.family="AlusX"
complement(2143..2155)
/rpt.family="Alus9"
complement(2156..2194)
/rpt.family="(TAAAA)n"
complement(2195..2480)
/rpt.family="Alus9"
complement(2481..2615)
/rpt.family="AlusX"
complement(2635..2652)
/rpt.family="POLY_A"
2949..3260
/rpt.family="AluJb"
complement(3391..3401)
/rpt.family="Alus9"
complement(3402..3428)
/rpt.family="(TAAA)n"
complement(3429..3677)
/rpt.family="Alus9"
3714..3884
/rpt.family="FRAM"
4143..4219
/rpt.family="L2"
complement(5060..5364)
/rpt.family="Alusq"
6307..7148
/rpt.family="L2"
7161..7436
/rpt.family="L2"
7508..7647
/rpt.family="L2"
complement(9623..9918)
/rpt.family="AlusX"
10006..10304
/rpt.family="Alusql"
10312..10478
/rpt.family="Alusq"
10508..10530
/rpt.family="AT-rich"
complement(15332..15839)
/rpt.family="AlusX"
complement(15844..16126)
/rpt.family="AlusX"
17035..17334
/rpt.family="AlusX"
17423..17720
/rpt.family="AlusX"
complement(18239..18513)
/rpt.family="AluJo"
18587..18814
/rpt.family="MIR"
complement(20103..20260)
/rpt.family="L1ME3A"
complement(20340..20494)
/rpt.family="AluJb"
complement(20495..20793)
/rpt.family="AlusX"
complement(20794..20934)

```



```

location/VuGall121
1..3626
/organism="Mus sp."
/strain="129SV"
/db_xref="taxon:10095"
/cell_type="ES cell"
1..1530
1531..1798
/number=1
1799..2240
/number=1
2241..2393
/number=2
2394..2485
/number=2
2486..3547
/number=3
2629..>3546
/codon_start=1
/product="oxytocin receptor"
/protein_id="BAAL18996.1"
/db_xref="GI:1902966"
/translation="MEGTAAANWSTIEDLGGVPGAEGNLTAGPPRNEALARVEA
VICILFLALSGNVCILALRTTRKHKSRLFFPKHSIADLVAVQVLPQLMDVT
FRFYGPDLICRLVKYFLOVGMFASYYLLMLLSGDLACIQPLASLRRTDLRAVL
WGLGCVSPVQVHHFYSRLREVAGVDFDCWAFIQPMGPKAYVTWTLAVYIVETVLAA
CYGLISFKYIWNLRKLTAAAGAAEGSDAAGGAGRAALARVSSVKLISKAKITRTVK
FIIVLIVGWTPPEFFYQMKWSVDVNAPKE"
3548..3626
/number=3

```

BASE COUNT	782 a	921 c	1131 g	792 t
Query Match	14.5%	Score 34.2:	DB 12:	Length 3626;
Best Local Similarity	60.08;	Pred. No. 2.8;	38;	Indels 0; Gaps 0;
Matches	57;	Conservative	0;	Mismatches
Qy	106	ACAAATCTCTTATGCTGACGCTCTGCTGAATGCTTGAAGCTGCTGCTCTCTCTC 165		
Db	2085	AAAGATTTACTAGGGTACATCTCTCTGGGTCCGCCCAAGCCATCTCTCTCAC 2026		
Qy	166	CATCTCCCTTCAGGACACGGTCACTCCACCA 200		
Db	2025	CTCTGCTTCAGAGACTCAGCCAGCTCCGCA 1991		
RESULT 7				
AC012184/c				
LOCUS	AC012184	65673 bp	DNA	HTG 21-OCT-1999
DEFINITION	Homo sapiens chromosome 16 clone RPCL-11_529K1, *** SEQUENCING IN PROGRESS ***, 40 unordered pieces.			
ACCESSION	AC012184			
VERSION	AC012184.1	GI:5091684		
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 65673)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Sequencing of Human Chromosome 16			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 65673)			
AUTHORS	DOE Joint Genome Institute.			
TITLE	Direct Submission			
JOURNAL	Submitted (21-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			
COMMENT	www.jgi.doe.gov.			
	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 40 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Gaps between the contigs are represented as			
	* runs of N, but the exact sizes of the gaps are unknown.			
	* This record will be updated with the finished sequence			
	* as soon as it is available and the accession number will			
	* be preserved.			
	* 1 760: contig of 760 bp in length			
	* 761 gap of unknown length			
	* 1560: contig of 800 bp in length			
	* 1561 gap of unknown length			
	* 2303: contig of 743 bp in length			
	* 2304 gap of unknown length			
	* 3255: contig of 952 bp in length			
	* 3256 gap of unknown length			
	* 4246: contig of 991 bp in length			
	* 4247 gap of unknown length			
	* 5603: contig of 1357 bp in length			
	* 5604 gap of unknown length			
	* 6762: contig of 1159 bp in length			
	* 6763 gap of unknown length			
	* 7004: contig of 242 bp in length			
	* 8046: contig of 1042 bp in length			
	* 8047 gap of unknown length			
	* 9309: contig of 1263 bp in length			
	* 9310 gap of unknown length			
	* 10988: contig of 1679 bp in length			
	* 10989 gap of unknown length			
	* 12037: contig of 1049 bp in length			
	* 12038 gap of unknown length			
	* 13400: contig of 1363 bp in length			
	* gap of unknown length			

```

* 13401 14099: contig of 699 bp in length
* 14100 15399: contig of 1300 bp in length
* 15400 16772: contig of 1373 bp in length
* 16773 17868: contig of 1096 bp in length
* 17869 18643: contig of 775 bp in length
* 18644 19788: contig of 1145 bp in length
* 19789 21845: contig of 2057 bp in length
* 21846 23223: contig of 1378 bp in length
* 23224 25010: contig of 1787 bp in length
* 25011 25466: contig of 456 bp in length
* 25467 27098: contig of 1632 bp in length
* 27099 28274: contig of 1176 bp in length
* 28275 29748: contig of 1474 bp in length
* 29749 31363: contig of 1615 bp in length
* 31364 33150: contig of 1787 bp in length
* 33151 36067: contig of 2917 bp in length
* 36068 37581: contig of 1514 bp in length
* 37582 39288: contig of 1707 bp in length
* 39289 42139: contig of 2851 bp in length
* 42140 44809: contig of 2670 bp in length
* 44810 47022: contig of 2213 bp in length
* 47023 51524: contig of 4502 bp in length
* 51525 53246: contig of 1722 bp in length
* 53247 56192: contig of 2946 bp in length
* 56193 58896: contig of 2704 bp in length
* 58897 61546: contig of 2650 bp in length
* 61547 65673: contig of 4127 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..65673
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="16"
                /clone="RPCI-11.529k1"
BASE COUNT 16932 a 16632 c 16336 g 15686 t 87 others
ORIGIN

```

```

Query Match      14.38; Score 33.4; DB 42; Length 65673;
Best Local Similarity 60.48; Pred. No. 5.8;
Matches 55; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 144 TGAAGGCTGCTCTCTCCATCTCCATCTCCCTTCAGGACACGTCACCTCCACCATGC 203
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
39815 TGAAGAACTCTCATCCACCTCTCGTCTCCCTCTTGAGCGGGGACCCCTGACCCCTC 39756
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 AAGATCTCAACACCATGTTGTCTGCAACACA 234
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 39755 TAAACTTCAAACCGCAAAATTCCTGTGTGCA 39725

RESULT 8
HS198C21

LOCUS Human DNA sequence from clone 198C21 on chromosome Xq26.1-26.3
DEFINITION Contains GPC4 (glypican 4), ESTs, STSS and GSS, complete sequence.

ACCESSION AL034400
VERSION AL034400.2 GI:4455461

KEYWORDS HTG; GPC4.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Hinxton, Cambridge; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.

TITLE 1 (bases 1 to 141762)

JOURNAL Heath, P.

Direct Submission
Submitted (16-MAR-1999) Sanger Centre, Hinxton, Cambridge; Sanger,
CS10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

requests: clonerequest@sanger.ac.uk

On Mar 21, 1999 this sequence version replaced gi:3980345.

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission

corresponding to the overlapping clone, as we submit sequences with

only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 198C21.

It may be shorter because we only sequence overlapping sections

once, or longer because we arrange for a small overlap between

neighbouring submissions.

The true right end of clone DJ358H7 is at 100 in this sequence.

This sequence has been finished according to sequence map criteria

as follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

This sequence was generated from part of bacterial clone contigs of

human chromosome X, constructed by the Sanger Centre Chromosome X

Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/ChrX

198C21 is from the library RPC16 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see http://bacpac.med.buffalo.edu/VECTOR: pPAC4.

Location/Qualifiers
1..141762

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/map="q26.1-26.3"

/clone="RP6-198C21"

/clone_lib="RPCI-6"

repeat_region

1..397

/note="Mt1D repeat: matches 100..500 of consensus"

repeat_region

757..810

/note="27 copies 2 mer to 85% conserved"

repeat_region

767..810

/note="11 copies 4 mer tctc 93% conserved"

repeat_region

813..1075

/note="AluJB repeat: matches 1..274 of consensus"

repeat_region

1186..1304

/note="THE1B repeat: matches 326..364 of consensus"

repeat_region

1205..1504

/note="THE1B repeat: matches 1..301 of consensus"

repeat_region

1505..1830

/note="THE1B repeat: matches 1..326 of consensus"

repeat_region

1974..2325

/note="match: 267843 STS containing (CA) repeat"

misc_feature

2115..2143

/note="14 copies 2 mer ca 100% conserved; differs from

267843"


```

RESULT 9
AC005090/c
LOCUS AC005090 116218 bp DNA PRI 27-AUG-1998
DEFINITION Homo sapiens BAC clone RG315L10 from 7p14-p15, complete sequence.
ACCESSION AC005090
VERSION AC005090.1 GI:3478661
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 116218)
AUTHORS Dauphin, S. and Blewald, T.
TITLE The sequence of Homo sapiens BAC clone RG315L10
JOURNAL Unpublished (1998)
REFERENCE 2 (bases 1 to 116218)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 3 (bases 1 to 116218)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (27-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Aug 27, 1998 this sequence version replaced gi:3212901.
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone; and the assembly was
confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and
sequencing collaboration between the NHGRI Chromosome 7 Mapping
Project (Eric D. Green, Director), John D. McPherson in the
Department of Genetics (Washington University), and the Washington
University Genome Sequencing Center. For additional information
about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
Clone RG315L10 is from a release of the human BAC library
CITB-HS-A. The library contains cloned DNA from human sperm. See:
Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
Kim et al., Genomics 34:213-8 (1996). The clone is available from
Research Genetics, Inc. (http://www.resgen.com).
VECTOR: pBelobAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RG315L10;
actual end is at 116218 of RG315L10.

Bases 41827 to 41835 of this sequence are single stranded. Base
41828 is of low quality and could be a G or an A.
Location/Qualifiers

source
1..116218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7p14-p15"
/clone_lib="CITB-HS-A"
/clone="RG315L10"
540..718
/rpt_family="MER1_type"
749..918
/rpt_family="MIR"
1003..1053
/rpt_family="L2"
2218..2531
/rpt_family="MaLR"
2900..3105
/rpt_family="MIR"
3554..3647
/rpt_family="L2"
4059..4294
/rpt_family="Alu"
4408..4462
/rpt_family="L1"
4503..5056
/rpt_family="L1"
5064..5576
/rpt_family="L1"
5566..6238
/rpt_family="L1"
7617..8092
/rpt_family="MaLR"
8150..8385
/rpt_family="Retroviral"
8629..9197
/rpt_family="Retroviral"
10415..10644
/rpt_family="Alu"
10648..10706
/rpt_family="L1"
12134..12568
/rpt_family="L2"
13525..13593
/rpt_family="MIR"
14875..15008
/rpt_family="L2"
15013..16289
/rpt_family="Mariner"
16362..16519
/rpt_family="MIR"
16760..17051
/rpt_family="L2"
17675..18091
/rpt_family="MaLR"
18474..18783
/rpt_family="L1"
18790..19571
/rpt_family="L1"
19567..24804
/rpt_family="L1"
24822..25147
/rpt_family="MER4-group"
25141..25312
/rpt_family="MER4-group"
25315..25572
/db_xref="GI:388822"
26886..27047
/rpt_family="MIR"
27462..27574
/rpt_family="MER81"
27695..27995
/rpt_family="Alu"
28050..28155
/rpt_family="MER21_g"

```

repeat_region	28124..28262	/rpt_family="MER21_g"
repeat_region	28335..28657	/rpt_family="MER21_g"
repeat_region	28927..28969	/rpt_family="L2"
repeat_region	30119..30461	/rpt_family="MaLR"
repeat_region	30476..30604	/rpt_family="MIR"
repeat_region	30989..31141	/rpt_family="MIR"
repeat_region	31207..31482	/rpt_family="Alu"
repeat_region	31827..31721	/rpt_family="MIR"
repeat_region	31855..31927	/rpt_family="MIR"
repeat_region	32385..32584	/rpt_family="L2"
repeat_region	32740..33139	/rpt_family="Retroviral"
repeat_region	33244..33314	/rpt_family="MIR"
repeat_region	33315..33621	/rpt_family="Alu"
repeat_region	33622..33745	/rpt_family="MIR"
repeat_region	34116..34418	/rpt_family="Alu"
repeat_region	36139..36443	/rpt_family="Alu"
repeat_region	36572..37062	/rpt_family="L2"
repeat_region	37202..37316	/rpt_family="MER1_type?"
repeat_region	37836..37899	/rpt_family="MER1_type?"
repeat_region	38128..38372	/rpt_family="L2"
repeat_region	39143..39335	/rpt_family="MIR"
repeat_region	39481..39628	/rpt_family="MIR"
repeat_region	39846..39890	/rpt_family="L2"
repeat_region	39917..40378	/rpt_family="L1"
repeat_region	40379..41802	/rpt_family="L1"
repeat_region	41853..41939	/rpt_family="L1"
repeat_region	41949..42017	/rpt_family="L2"

Query Match	14.1%;	Score 33;	DB 11;	Length 116218;	
Best Local Similarity	51.0%;	Pred. No. 8;			
Matches	78;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0;
Y	74	CTTTGCAACTGAAGTCTTTCTGACAAATTCCTCTATGAGTCCAGGCTTCC	133		
b	81977	CTCTGTGTAGATGAAGCTAAATCTGACACTTTGGTCTTCGCTCTAAACCAT	81918		
Y	134	TGGAATGCTTGAAGCTCTGCT			

SOURCE Homo sapiens DNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (sites)
 Shina S., Tamiya G., Oka A. and Inoko H.
 Homo sapiens 2 229 818bp genomic DNA of 6p21.3 HLA class I region
 Published Only in Database (1999) In press
 2 (bases 1 to 100000)
 Hirakawa, H., Imanouchi, H., Imai, K. and Shimada, J.
 Direct Submission
 Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases, Mika
 Hirakawa, Japan Science and Technology Corporation (JST), Advanced
 Databases Department, 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
 Japan (E-mail: mika@tokyo.jst.go.jp).
 URL: http://www-alis.tokyo.jst.go.jp/, Tel: 81-3-5214-8491,
 Fax: 81-3-5214-8470)
 This sequence is conducted by Tokai University as a JST sequencing
 Team.
 Principal Investigator: Hidetoshi Inoko Ph.D
 Phone: +81-463-93-1121, Fax: +81-463-94-8884,
 The sequence is submitted by Human Genome Sequencing in ALIS
 project of JST
 Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
 For further information about this sequences, please visit our
 sequence archive web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
 html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
 Location/Qualifiers
 1..100000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p21.3"
 /complement(25387..25469)
 /note="316A/316S: The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:215643"
 STS
 /complement(36020..36262)
 /note="TIGR-A007C28: The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:4588360"
 STS
 /complement(36173..37150)
 /note="SGC34051: The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:4575750"
 STS
 /complement(41454..41670)
 /note="stSG3452: The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:4573047"
 STS
 56458..56539
 /note="960R: The location is between each flanking site of
 PCR primers."
 /db_xref="GDB:1297691"
 STS
 56463..56581
 /standard_name="D6S2266"
 /note="sv960h11-8: The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:5583962"
 STS
 80448..80597
 /note="SGC32967: The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:6456157"
 STS
 /complement(36072..36222)
 /note="WI-17346: The location is between each flanking site
 of PCR primers."
 /db_xref="GDB:4580894"
 STS
 /complement(96558..96659)
 /note="2185/2186: The location is between each flanking
 site of PCR primers."
 /db_xref="GDB:581290"
 STS
 25730 a 23716 c 23917 g 26637 t
 "GIN"

Query Match 14.0%; Score 32.8; DB 10; Length 100000;
 Best Local Similarity 59.0%; Pred. No. 9.1; Indels 0; Gaps 0;
 Matches 58; Conservative 0; Mismatches 42;
 QY 64 CGGCTATGCTTCGCACTGAGCTGAGAGCTCTTTCTGACAAATTCCTCTATGAG 123
 Db 91709 CGGCTATGCTTCGCACTGAGCTGAGAGCTCTTTCTGACAAATTCCTCTATGAG 123
 QY 124 TCCAGCTTCTGGAATGCTGAAAGCTCTGCTCTCTCC 163
 Db 91649 CCTAGATTACTGAATAGCCTCACTGCTCTCTCTCTCTCC 91610
 RESULT 12
 AC004209/c 43600 bp DNA PRI 03-SEP-1998
 LOCUS Homo sapiens clone UWGC:y55c025 from 6p21, complete sequence.
 DEFINITION AC004209
 ACCESSION AC004209
 VERSION AC004209.1 GI:3522963
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 43600)
 Janer M., Guillaudoux, T., Vu, Q., Kutyavin, T., Harter, H. and
 Geraghty, D.E.
 Large scale sequence analysis of the human MHC class I region
 Unpublished (1998)
 The Hutchinson Cancer Research Center
 The Clinical Research Division
 1100 Fairview Ave. N., P.O. Box 19024
 Seattle, WA 98109-1024
 2 (bases 1 to 43600)
 Geraghty, D.E. and Olson, M.V.
 Direct Submission
 Submitted (23-FEB-1998) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 43600)
 Geraghty, D.E. and Olson, M.V.
 Direct Submission
 Submitted (03-SEP-1998) Human Genome Center, University of
 Washington, Box 352145, Seattle, WA 98195, USA
 Fred Hutchinson Cancer Research Center
 The Clinical Research Division
 1100 Fairview Ave. N., P.O. Box 19024
 Seattle, WA 98109-1024
 Contact: Daniel E. Geraghty (geraghty@fhcr.org)
 On Sep 3, 1998 this sequence version replaced gi:2905851.
 COMMENT Overlapping Sequences:
 5': UWGC:y55c174
 3': UWGC:y55c068

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 Double stranded (DS) coverage: 79.5%
 DS or two chemistry coverage: 99.7%
 Single stranded regions: 2

 Sequence Validation:
 This sequence has been validated by Multiple Complete Digest
 Mapping. Comparison of the experimentally derived map digest
 fragments with sequence-predicted fragments is given below.

Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	HindIII	Map	Seq
590.87	5842.00	1561.21	1547.00	2899.98 2857.00
1246.59	1240.00	1365.20	1360.00	4696.59 4486.00
628.61	627.00	1803.37	1799.00	8400.84 8055.00
12412.85	11969.00	13189.91	12685.00	10629.14 10329.00
9189.25	8932.00	4298.12	4138.00	951.62 941.00
2647.86	2596.00	736.47	727.00	7859.84 7632.00
989.83	992.00	2883.77	2832.00	
988.83	977.00	1210.44	1211.00	
6670.61	6509.00	8505.16	8204.00	
		7741.77	7605.00	

FEATURES

Source	Location/Qualifiers
1.	43600
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="6"	
/clone="CGM1:DL23C12"	
/sub_clone="UMGC:Y55C025"	
/cell_line="CGM1"	
/clone_lib="Wash U YAC Library"	
/map="6p21"	
repeat_region	complement(21..277)
repeat_region	/rpt_family="Alu"
repeat_region	676..873
repeat_region	/rpt_family="Alu"
repeat_region	complement(1150..1448)
repeat_region	/rpt_family="Alu"
repeat_region	2056..2345
repeat_region	/rpt_family="Alu"
repeat_region	complement(2630..2913)
repeat_region	/rpt_family="Alu"
repeat_region	complement(6636..6708)
repeat_region	/rpt_family="Alu"
repeat_region	complement(6854..7129)
repeat_region	/rpt_family="Alu"
repeat_region	complement(7265..7543)
repeat_region	/rpt_family="Alu"
repeat_region	complement(7635..7770)
repeat_region	/rpt_family="Alu"
repeat_region	complement(8253..8535)
repeat_region	/rpt_family="Alu"
repeat_region	complement(8973..9090)
repeat_region	/rpt_family="Alu"
repeat_region	15507..15767
repeat_region	/rpt_family="Alu"
repeat_region	17528..17816
repeat_region	/rpt_family="Alu"
repeat_region	18354..18646
repeat_region	/rpt_family="Alu"
repeat_region	complement(19460..19743)
repeat_region	/rpt_family="Alu"
repeat_region	complement(26422..26590)
repeat_region	/rpt_family="Alu"
repeat_region	complement(26722..27017)
repeat_region	/rpt_family="Alu"
repeat_region	complement(27496..27795)

repeat_region	/rpt_family="Alu"	complement(27976..28261)
repeat_region	/rpt_family="Alu"	complement(28394..28664)
repeat_region	28716..28998	
repeat_region	/rpt_family="Alu"	complement(29874..29952)
repeat_region	/rpt_family="Alu"	complement(30110..30353)
repeat_region	complement(30451..30660)	
repeat_region	/rpt_family="Alu"	complement(30685..30865)
repeat_region	/rpt_family="MER42"	complement(30862..31150)
repeat_region	/rpt_family="Alu"	complement(31182..31468)
repeat_region	/rpt_family="Alu"	complement(31482..31678)
repeat_region	/rpt_family="MER42"	31795..32069
repeat_region	/rpt_family="Alu"	32112..32243
repeat_region	/rpt_family="Alu"	complement(32820..33022)
repeat_region	/rpt_family="MLT2B2"	33228..33415
repeat_region	/rpt_family="MLT2B2"	33680..33971
repeat_region	/rpt_family="Alu"	34331..34939
repeat_region	/rpt_family="Alu"	35200..35499
repeat_region	/rpt_family="Alu"	35908..36190
repeat_region	/rpt_family="Alu"	complement(36377..36494)
repeat_region	/rpt_family="LTR8"	36578..36830
repeat_region	/rpt_family="Alu"	37011..37330
repeat_region	/rpt_family="Alu"	complement(37530..37676)
repeat_region	/rpt_family="LTR8"	complement(37982..38276)
repeat_region	/rpt_family="Alu"	complement(38785..39164)
repeat_region	/rpt_family="Alu"	complement(39725..40004)
repeat_region	/rpt_family="Alu"	complement(40229..40521)
repeat_region	/rpt_family="Alu"	complement(40667..43600)
repeat_region	/rpt_family="L1"	

BASE COUNT 11071 a 5661 c 10338 g 12530 t
ORIGIN

Query Match 14.08; Score 32.8; DB 11; Length 43600;

Best Local Similarity 58.08; Pred.No.8.7; Mismatches 0; Indels 0; Gaps 0;

Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 64 CCGGTATGACTTGGCAACTGAAGAGTCTTTTCTGACAAATCTCTCTATGAG 123

Db 17165 CCGGCCCTGTCTCACCACCTCCAGATCTCTACTCTAGCAACTCTTATTTCT 17106

Oy 124 TCAGCTCTCTGGAATTCCTTTGAAAAGCTCTGCTCTCTCC 163

Db 17105 CCTAGATTACTGAAATAGCCTCAACTGCTCTCTCTGCTCC 17066

RESULT 13

```

AC002327      AC002327      159515 bp      DNA
DEFINITION    Mus musculus chromosome 7, clone 19K5, complete sequence.
ACCESSION     AC002327
VERSION       AC002327.1 GI:4210518
KEYWORDS      HTG.
SOURCE        house mouse.
ORGANISM      Mus musculus.

REFERENCE
AUTHORS       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE         1 (bases 1 to 159515)
JOURNAL       Birren, B., Linton, L., Nusbaum, C. and Lander, E.
              Mus musculus chromosome 7, clone 19K5
              Unpublished
REFERENCE     2 (bases 1 to 159515)
              Hawkins, T.L., Reeve, M.P., Christoffersen, A., Birren, B.W.,
              Fasman, K.H., Lander, E.S., McKernan, K., Munro, C., Richardson, P.,
              Barna, N., Chang, A., Cooke, P., Daly, M.J., Devon, K., Dewar, K.,
              Forrest, C., Gage, D., Geraigery, K., Guitau, G., Hagos, B., Huang, J.,
              Jacotot, L., Lane, M., Lee, K., MacKenzie, J., Marquis, N.,
              McDermott, J., Molla, M., Moloney, N., Morrow, J., Nachman, A.,
              Taylor, J., O'Connor, T., Olotu, A., Peterson, K., Rollins, G.,
              Spencer, J., Stilwell, J., Stone, C., Strickland, C., Sydney, K.,
              Traish, A., Wilmer, F., Zemtseva, I. and Zody, M.
              Direct Submission
              Submitted (15-JUL-1997) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              3 (bases 1 to 159515)
              Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
              Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
              Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,
              Cooke, P., DeArelano, K., Depayre, E., Devon, K., Dewar, K.,
              Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
              Funk, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
              Hagos, B., Heaford, A., Horton, L., Howland, J.C., Jones, C., Kann, L.,
              Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
              Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
              Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
              Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
              Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
              Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
              Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
              Wheeler, J., Wu, X., Wyman, D., Ye, W.J. and Zody, M.
              Direct Submission
              Submitted (02-FEB-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Charles Street, Cambridge, MA 02141, USA
              On Feb 2, 1999 this sequence version replaced gi:4204708.
              All repeats were identified using RepeatMasker: Smit, A.F.A. &
              Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source
1..159515
   Location/Qualifiers
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="19K5"
   /clone_lib="Research Genetics/Cal Tech CITB978SK-B (plates
   1-194)"
   /map="7"
   /chromosome="7"
   complement(81..193)
   /rpt_family="AluOf/FLAM"
   718..860
   /rpt_family="TA)n"
   863..942
   /rpt_family="AT_rich"
   1099..1151
   /rpt_family="AT_rich"
   1482..1542
   /rpt_family="FRAM/FAM"
   1753..1782
   /rpt_family="AT_rich"
   1908..1957
   /rpt_family="Alu"

repeat_region 2052..2128
               /rpt_family="Alusg/x"
repeat_region 2709..2759
               /rpt_family="Alusg"
repeat_region 2806..2832
               /rpt_family="(A)n"
repeat_region 2852..2914
               /rpt_family="MIR"
repeat_region 2996..3022
               /rpt_family="AT_rich"
repeat_region 3827..3913
               /rpt_family="L2"
repeat_region 4603..4679
               /rpt_family="(CGG)n"
repeat_region 4722..4817
               /rpt_family="(CGG)n"
repeat_region 4832..4954
               /rpt_family="GC_rich"
repeat_region 6796..6818
               /rpt_family="(TTCC)n"
repeat_region 6964..7138
               /rpt_family="MERSA"
repeat_region 7154..7191
               /rpt_family="(TC)n"
repeat_region complement(8423..8550)
               /rpt_family="MIR"
repeat_region 8635..8751
               /rpt_family="MIR"
repeat_region complement(9506..9667)
               /rpt_family="Alusg/x"
repeat_region complement(9699..9812)
               /rpt_family="L2"
repeat_region 10967..11107
               /rpt_family="FRAM"
repeat_region 11756..11822
               /rpt_family="(TTTTA)n"
repeat_region complement(11874..11966)
               /rpt_family="Alusg"
repeat_region 12017..12047
               /rpt_family="(TTTTA)n"
repeat_region complement(12138..12196)
               /rpt_family="RNA-Lys-AAA"
repeat_region 12402..12515
               /rpt_family="MIR"
repeat_region 12665..12749
               /rpt_family="FLAM_A"
repeat_region complement(12899..12968)
               /rpt_family="MIR"
repeat_region 13967..14023
               /rpt_family="(TCCC)n"
repeat_region 14765..14811
               /rpt_family="AT_rich"
repeat_region complement(15408..15604)
               /rpt_family="MES102"
repeat_region complement(16028..16174)
               /rpt_family="FLAM_C"
repeat_region 18347..18468
               /rpt_family="MIR"
repeat_region complement(18522..18577)
               /rpt_family="MIR"
repeat_region complement(19193..19405)
               /rpt_family="AluJb"
repeat_region 19549..19601
               /rpt_family="PNERI"
repeat_region 23117..23147
               /rpt_family="AT_rich"
repeat_region complement(23988..24046)
               /rpt_family="RNA-Ala-GCG"
repeat_region 24457..24544
               /rpt_family="FLAM_A"
repeat_region 25145..25204
               /rpt_family="MIR"
repeat_region 25289..25322

```

```

repeat_region      /rpt_family="AT-rich"
                    complement(25541..25624)
repeat_region      /rpt_family="L2"
                    25694..25743
repeat_region      /rpt_family="(TTCA)n"
                    26325..26466
repeat_region      /rpt_family="AluSg"
                    27277..27299
repeat_region      /rpt_family="(TTTG)n"
                    complement(27302..27443)
repeat_region      /rpt_family="FLAM_C"
                    complement(27608..27662)
repeat_region      /rpt_family="MIR"
                    complement(28409..28532)
repeat_region      /rpt_family="MIR"
                    complement(30322..30465)
repeat_region      /rpt_family="L2"
                    35588..35800
repeat_region      /rpt_family="AluJ"
                    40654..40703
repeat_region      /rpt_family="A-rich"
                    42215..42257
repeat_region      /rpt_family="CT-rich"
                    43511..43544
repeat_region      /rpt_family="AT-rich"
                    43715..43742
repeat_region      /rpt_family="(T)n"
                    43745..43767
repeat_region      /rpt_family="AT-rich"
                    43817..43956
repeat_region      /rpt_family="AluJ"
                    44196..44221
repeat_region      /rpt_family="(CCCTAA)n"
                    44395..44471
repeat_region      /rpt_family="AluSp"
                    46141..46271
repeat_region      /rpt_family="FRAM"
                    46603..46641
repeat_region      /rpt_family="AT-rich"
                    47045..47073
repeat_region      /rpt_family="(CA)n"
                    47293..47357
repeat_region      /rpt_family="(GA)n"
                    complement(48842..49005)
repeat_region      /rpt_family="FRAM"
                    49023..49127
repeat_region      /rpt_family="AluSp"
                    49498..49629
repeat_region      /rpt_family="FLAM_C"
                    complement(50407..50479)
repeat_region      /rpt_family="L2"
                    50860..50920
repeat_region      /rpt_family="AluS"
                    50929..50965
repeat_region      /rpt_family="(GTTG)n"

Query Match      14.0% Score 32.8; DB 12; Length 159515;
Best Local Similarity 53.0%; Pred. NO. 9.4;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 99 TTTTCGACAAATCTCCCTATGATGCTCAGCTTCCTGGGARTGCTTGAAAGCTCTGCCT 158
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52401 TTTTACTGTTTATTTTAAATGAGCCCACTTCTATGATGACCAAGGATGACTGACCT 52460
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 159 CCTCTCCATCTCCCTTCAGGAGCAGGCTCACCCTCCACCACCAAGATCTCAACACCA 218
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 52461 CCTGTCCACCTCTCAGTTGGGAGTACGGTCAGTCAGTACCCCATCTCTATCTCTACGCT 52520
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 219 TGTGTCTGCAA 230
    ||| ||| |||
Db 52521 GGCTCTATCA 52532
    ||| ||| |||

```

RESULT 14 Hs179N16/c

LOCUS
DEFINITION

Hs179N16 172048 bp DNA PRI 23-NOV-1999
Homo sapiens DNA sequence from PAC 179N16 on chromosome
p21.1-21.33. Contains the SAPK4 (MAPK p38delta) gene, and the
alternatively spliced SAPK2 gene coding for CSaids binding protein
CSBP2 and a MAPK p38beta LIKE protein. Contains ESTs, STSs and two
predicted CpG islands, complete sequence.

ACCESSION

VERSION

KEYWORDS

295152 1 GI:3036773
HNG: CpG island; CSaids binding protein; CSBP2; MAPK; Mitogen
activated protein kinase; MXI2; P38; SAPK4; Stress activated
protein kinase.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172048)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (02-APR-1998) Chromosome 6 Project Group
(http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquies@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On Apr 8, 1998 this sequence version replaced gi:2505919.

IMPORTANT: This sequence is the entire insert of clone 179N16.

During sequence assembly data are compared from overlapping clones.

Where differences are found these are annotated as variations

together with a note of the overlapping clone name. Note that the

variations annotated may not be found in the sequence submission

corresponding to the overlapping clone as we submit sequences with

only a small overlap as described above.

This sequence was generated from part of bacterial clone contigs of

human chromosome 6, constructed in collaboration by the Sanger

Centre chromosome 6 mapping group and Ioannis Ragousis. Further

information can be found at http://www.sanger.ac.uk/HGP/Chr6/ This

sequence has been finished according to sequence map criteria as

follows. An attempt is made to resolve all sequencing problems,

such as compressions and repeats, but not necessarily within known

annotated human repeat sequence elements (e.g. Alu). Where the

sequence is ambiguous, there is an annotation using the 'unsure'

feature key.

The true left end of clone 179N16 is at 1 in this sequence. The

true right end of clone 524E15 is at 690.

The true left end of clone 179N16 is at 172048.

179N16 is from the library RPI1 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong.

For further details see http://bacpac.med.buffalo.edu/.

FEATURES

source

1..172048

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="p21.1-21.33"

/clone="RPI-179N16"

/clone_lib="RPI-1"

complement(1..1077)

/note="match: STS ALU020972"

4..457

/note="11 repeat: matches 5082..4624 of consensus"

493..734

/note="AluJo repeat: matches 302..62 of consensus;

incomplete repeat"

1714..1954

/note="LIMB8 repeat: matches 324..574 of consensus"

1956..2257

/note="AluJ repeat: matches 1..302 of consensus"

2258..2543

/note="LIMB6 repeat: matches 566..874 of consensus"

2548..2848

/note="AluSx repeat: matches 301..1 of consensus"

2879..3160

/note="AluJo repeat: matches 277..2 of consensus;

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region incomplete repeat"
3168..3369 /note="MER44C repeat: matches 258..463 of consensus"
repeat_region 4208..4497 /note="Aluub repeat: matches 289..1 of consensus"
repeat_region 5473..5894 /note="MLTRIC repeat: matches 39..465 of consensus"
repeat_region 5986..6284 /note="Aluub repeat: matches 1..296 of consensus"
repeat_region 6361..6420 /note="AluSx repeat: matches 301..242 of consensus;
incomplete repeat"
6512..6914 /note="Aluub repeat: matches 2..302 of consensus"
repeat_region 6951..7084 /note="FLAM_C repeat: matches 1..133 of consensus"
repeat_region 7621..7941 /note="L1 repeat: matches 5068..5387 of consensus"
repeat_region 7809..8170 /note="L1MC2 repeat: matches 15..384 of consensus"
repeat_region 8255..8378 /note="MER2 repeat: matches 13..136 of consensus"
repeat_region 8396..8603 /note="Aluub repeat: matches 291..82 of consensus;
incomplete repeat"
8605..8691 /note="MER2 repeat: matches 125..209 of consensus"
repeat_region 8826..8890 /note="MER2 repeat: matches 279..344 of consensus"
repeat_region 9518..9894 /note="THE1C repeat: matches 3..371 of consensus"
repeat_region 9964..9989 /note="13 copies of 2 mer 96 % conserved"
repeat_region 10970..11269 /note="AluSg repeat: matches 300..1 of consensus"
repeat_region 12207..12333 /note="MIR repeat: matches 106..248 of consensus"
repeat_region 12735..13031 /note="AluSx repeat: matches 1..298 of consensus"
repeat_region 13042..13256 /note="AluSg repeat: matches 2..217 of consensus;
incomplete repeat"
13494..13581 /note="AluSx repeat: matches 205..288 of consensus;
incomplete repeat"
13595..13953 /note="THE1B repeat: matches 364..1 of consensus"
repeat_region 13972..14648 /note="L1MC3 repeat: matches 1676..2345 of consensus"
repeat_region 14649..14948 /note="AluSx repeat: matches 1..301 of consensus"
repeat_region 14949..15077 /note="L1MC3 repeat: matches 2336..2470 of consensus"
repeat_region 15102..15231 /note="FLAM_A repeat: matches 133..4 of consensus"
repeat_region 15244..15289 /note="23 copies of 2 mer 100 % conserved"
misc_feature 15674..15853 /note="match: 263849 CpG Island clone 9ic11"
repeat_region 16341..16702 /note="MIR2FA repeat: matches 33..394 of consensus"
repeat_region 16823..16969 /note="MER11A repeat: matches 4..151 of consensus"
repeat_region 16968..17594 /note="MER11A repeat: matches 60..738 of consensus"
repeat_region 17278..17886 /note="MER11B repeat: matches 2..627 of consensus"
repeat_region 17977..18027 /note="MIR2FA repeat: matches 400..450 of consensus"
repeat_region 19217..19333 /note="MIR repeat: matches 262..117 of consensus"
repeat_region 19471..19693 /note="AluJo repeat: matches 85..301 of consensus;
```

Wed Mar 22 12:07:50 2000

DEFINITION	Equine herpesvirus 2, complete genome.	
ACCESSION	U20824	
VERSION	U20824.1	GI:695172
KEYWORDS		
SOURCE	Equine herpesvirus 2	
ORGANISM	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Betaherpesvirinae	
REFERENCE	1 (bases 1 to 184427)	
AUTHORS	Telford, E.A., Watson, M.S., Aird, H.C., Perry, J., and Davison, A.J.	
TITLE	The DNA sequence of equine herpesvirus 2	
JOURNAL	J. Mol. Biol. 249 (3), 520-528 (1995)	
MEDLINE	95302501	
REFERENCE	2 (bases 1 to 184427)	
AUTHORS	Telford, E.A.R.	
TITLE	Direct Submission	
JOURNAL	Submitted (09-FEB-1995) Elizabeth A.R. Telford, Institute of Virology, MRC Virology Unit, University of Glasgow, Church Street, Glasgow G11 5JR, United Kingdom	
FEATURES	Location/Qualifiers	
source	1..184427	
	/organism="Equine herpesvirus 2"	
	/strain="86/67"	
	/db_xref="taxon:12657"	
repeat_region	1..17553	
	/rpt_type="direct"	
CDS	9248..10399	
	/note="ORF E1"	
	/codon_start=1	
	/product="g protein-coupled receptor"	
	/protein_id="AAC13788.1"	
	/db_xref="GI:695173"	
	/translation="MATTSAITSTVNTSSLAATTMTFTSLTSTVTTIASLVPSTNSS IKNTDMDVDVDESPACPKYKSDTTRLAQVVPALYLLVFLFGLGNILVVIIRYMK TKNITMLLNLAIIDLFLTLFPMWYIGMTYHDTWTFGSLKLLRGVCYMSLYSQV CILLVDRYLAAYVYATLRFRTVTCGIVCTWFLAGLISLPEFFHGHQDDNG RVQGDYKPKMSNWRARHAKVIMLSLIPLLIMAYCVYVIRLLRRPKKKYKA IKLFIWYAVFYVFTPNVILLSTFHTLLNLOALSSNLDMLLITKTVAYTHCC INPVIYAFGEKERRLLHFFHTYVAIYLCIKYIFLSDGSGEGPTRI"	
	10494..10499	
polyA_signal	/note="ORF E1"	
	10660..10665	
polyA_signal	/note="ORF E1"	
	12533..13114	
repeat_region	14197..14541	
repeat_region	18810..19733	
CDS	/note="ORF E2; product has hydrophobic N-terminus"	
	/codon_start=1	
	/protein_id="AAC13789.1"	
	/db_xref="GI:695174"	
	/translation="MIILGLVLSLLASTASIAQAMPOVITGHEGRPLSVRCHVAGK NPSKPTMLLKTIEGSAFRGTIVPRFGPVYTKSSCGSGTGVTVDSLMDAGRYA CMVQNDPKDESQSDYHTATNTIVLETRRRSLGQPTAGRSQLSQITGP LSSQSESVVAVKGVVGVVGSLSLRECOFSGATAGLNSAVLFRKVGAGYQAVST ISTSLITLDSLSPVIEGTQAKWCGVNATKEDSGEYMCIMGDFPEGESEDGDFSG VDFEKITVTPEDGADDDLYLTSSEKK"	
	19761..19766	
polyA_signal	/note="ORF E2"	
	19784..19789	
polyA_signal	/note="ORF E2"	
	20320..20844	
CDS	/note="ORF E3; product has N- and C- hydrophobic domains and 6 N-linked glycosylation sites"	
	/codon_start=1	
	/protein_id="AAC13790.1"	
	/db_xref="GI:695175"	
	/translation="MATEVGFACALVILGGYAGTPTSKSTLYNSQNTVPSIE NGQSLSYNGSNFIRYEFNCSGIEQLQGWTKTCIFRNGFIWINGPSPCKLNTT TLMTSTSTPTVTGTFPDQNTHTHTVTKVPRINDRLNLRFGITPWAITLWILL LVWIVNCCGPMF"	
	21360..25319	
CDS	/note="ORF 03"	
	/codon_start=1	
	/product="tegument protein"	
	/protein_id="AAC13791.1"	
	/db_xref="GI:695176"	
	/translation="MDSTSKRSFRKARRGGRATVYFVPTWQHGGQGGQPVSSFAFSL VNEVGFGLFDYGCAGTGATAGRAAADVPVGVASEAFMEILRTVLPHGRVAGVSELP AGLSYLTSDDGAGGPGCTLSFDVLSHSIKQHAHYMLCNOLSSPASFCRTPGYLGFTPLG GGEATPPIKILAVFFHTLWSLHSAKQHAHYMLCNOLSSPASFCRTPGYLGFTPLG DPEGRDLKFAFVFPALAKADAKALEMFREHDSVLTGTHGCAIKIRPNENGL LGLYRTAPTSRVFIDRVFSSGKPDYFYVYVGGISGASELLOLLISEAGFEGM GNVGLHGVVGGSDGDFALQYLESHMLAQRCGAPTLGLFVRVMDGSAAPFGMGPA SPHFATGATRASIIISPKACISPVRGGSVFMGLGDFFORQAGGAYASHDSGRKLA SILAAVQTFINAGGRAVLTSCQSLGCTHAKIAAACGVMGLKTYTSLLEPARGSEL SHFSPGNAMNRSILSKFNLVEAVLLTVNTPNRQILGKFTGFCRVSLEHSL ABAKIYLDSDSPDYCQGPYLCFKMKPRYALPIDRMCTGYSTRPTSPVNS SDAGPDSAREVILGILRHTPGVCKNYIVKHVDRCSAGVACQCGVPLDLPVADYSM VYVESAGPSSMEPTWETROVDDCELIHEDGVSGVCSALGCKSVFPYFEACAK MAIASEVLSALAFIARIEDIVYVLSLAWHLKGAQGEVFTLLTRCKRFCEADGVSCN VTSCASSRRAGGDPPEGGADIKSFVASIAYKLAVALHLRAKRAVSGHDVGGGLM GKHAFAVYQVGRKQYTGSPVNSIGAYLGKARAVLHLRAKRAVSGHDVGGGLM LCVRVGTGKTCERFTRITRGDGMGLTSETPGVALEVPSSASAAYEILRSKIY AALAEAMAMNRSIDVEVPVLDGDMGLTSETPGVALEVPSSASAAYEILRSKIY EGPYERERAEFLAYTPFRLINRNKHRVALLPGCPYLAALDALDSGDFALAVS HLDLAHVLTVDEVCGLFVAGVDNTSDTKTSDALTSALVHSRNPVIRDFLSKDTFSLG VSMACRILFDGMAYSPALQSRDRCVPVAVCKSFESRMLNFINQSTPAVAFETLRG SLPCWAGQTHLFGAPSPCECLDRVAVSGQAASMFYGSVDVAPAMAYPANTETLPV AGICSDGRHVLJLHDITASYILMQVHPVPSKNIPIAVSPMKQVQFYDLHANVTASH"	
	25639..26190	
	/note="ORF E4"	
	/codon_start=1	
	/protein_id="AAC13792.1"	
	/db_xref="GI:695177"	
	/translation="MVPSEFFFEEDRVRLENEAQLSLTNLLKNVFAQTLDMPKPEGV LTTTEALLAWVDSCKKEYLHOLIELVMQVPSVEAPTSAINSLGIRTHQGGSD NFGRILCSISFSCFELVLQSDERCLSVFSELAKEYVESQNLWLAISGGLSAGLRE RFRSLVFPALAKQKWLRF1FFK"	
	26436..29873	
	/note="ORF 06"	
	/codon_start=1	
	/product="single-stranded DNA binding protein"	
	/protein_id="AAC13793.1"	
	/db_xref="GI:695178"	
	/translation="MNSNRAAGPVEENVGTQASVGCPEVYLYPGDTFFVEEASLIG NLHAGEVFLPLSLGLTVAEDFHVNVKAVHKLPDPAIVSVKASVHREVIVFANAAC FKPIFAGPGLGCLCAASRQLGVAEPERRAGGAARPELADLGLLPGAESHIAGVVY TESFKERLYRGOLVYVESQIYGVGECDAFKVPLVDGELFAKSPCENRHYFYHAGV SRYLFEAHTLSLAQRALRVPGILGALERSFHQYKLPKVECEPPEATGHTGAGD CSITIYDVSATLAVSGLSFLVPOEGFALLSDKNPIEGECETPEORVEALQFNA QKAVHYHSQLFSGNSVLYLARVOKQASNRGGGENVNFSFMGHLACLAEPTQKENG YVGTAAITPWCLECGSPACVNTIFYRLDRFPVYVSVYRRDPYVTVGVAGATNDL DIAGNANYRDKDEESNOSERKFTYQVOTVTLERLSEAGICEGEGDVGDAHNIG SFYAFKEDIGLVDEVARFINSVKNVNYRESIKSIHHIVQYVNCVWQPPCPVEL NLYRVCVAVQDICIPLTCMMYKENPVGSPGEMLMKHVQTLWTNFKNSCIDKVL TGTIYKVVHKOCDFFEDVDASAGEVSVCKTQVRISRALMAPRVYKLNRIIFSNS TGTIYKVVHKOCDFFEDVDASAGEVSVCKTQVRISRALMAPRVYKLNRIIFSNS PFTFNPRESTVELANYVEQNSRLHGETSIIDVWPEFNYIAKVLNNAIFRACQTO FYATTIHLCTPIQTPAEYPHALGADVGEIILGAARELTYPVIVCTSRDICE VGMCRPIVTLPIVWNYKTVGTSQVAFQCANILGYIFGRVDKNIIPDAGFRKQVST SMKRHRVMTPLSDHLRRSVQGAFAIEGVRVRVQOILSDGDNPHVIRVQLV KSLGSECRSVSEYDLEYTMQYIFAGDVSERLQRLSDLGDSSEWALSVEGEDP LGSELEFEKVEDAECLGHPQDDEPALAPAAAPOYSGSSSVAGKRRKANVILGDL"	
	29997..32183	
	/note="ORF 07"	
	/codon_start=1	
	/function="role in capsid maturation/DNA packaging"	
	/protein_id="AAC13794.1"	
	/db_xref="GI:695179"	
	/translation="MARELAAYAAQVDFDLAAEYSLGGLYCDPSSIDKFCVMANSKVF KLCESLCLLQRLDTECPGLSLEQLHQLQNTREALVQTLDDLSGDSRSYEALHPS RLESGCKRHARVVPYFPGAEKTVLSLLNDVEYFVKRLANSFYCLPAEGALPCE RYVATLGRVSPIDPADAYVSSVPCASCFEAAMLNQGSEVLSMLAAYNCNVHCRCQ VPSDPVGVFENELRHLGADARAGRAGGREGREGRGDEDEDEEPRDQGDG GGGAAALRVLTSESLSLVAGHTIFEEEDGRLAELISNLVYSSNAADRGGVGGTARATSSS	

Job time: 1194 sec

HNKLFHEARHMRHRAWLGRGAPSHFFDAHRPSLESFCGVFNISIDDTIAALQKD
CSATFLKSNQTLIOQQNELYRLNEVLNAGRDEGGAKAEIAADAEPKPDGGAS
CPDRVLSDFQVRLDLYLKLRLDGLRRLTDCIETGRVLSDTLSLRWMSALASAA
RLNVEFRFQVGLGWADLTAGGEAFENSKYIKNALHGRLNREHLDSIVVHYRL
ITGPLSQNSHFVFPDPAVALAYCLDAAGAMPKOKLVITETWPGIESKDWIDCNFSF
YSIETGDLNLTQKTLNIREAVLSISLNRWKSLSLSLSATELRGSCLASASGEL
GGVYLVEGTAPLVLVDFSGYFKDLYLLYTHLQLSGRQASV
30114 .30119

polyA_signal

/note="orf E3, 03, E4 and 06"

CDS

32195 .34819

/note="orf 08"

/codon_start=1

/product="glycoprotein B"

/protein_id="AAC13795.1"

/db_xref="GI:695180"

/translation="MGVGGPVRVLCVVAALLCGVAQEVVAETTTFFATHRPEVV
AEENPAPLPPRVCGASPTGGEIREFPEESCPNTEKDHIEGIALIYKINIVYVF
NVRKYRIMSTIYKGSDEAIHQHRSYAVPLEVQMDHYIQCFSAVQVNEGHH
VNTYDRGNETAEIKPADGLTSITRQSQPEVATPRNLLMSYTRTYNCEVTE
MSARSKPFEFFVTSVGDTEIEMSPFLKENGTEPEKILKPHSIQLLKNTAVTKYVGL
QADNRTRFAIFGDSLSWKAITENSISYCDLLKMGFSNAIQTHNSLHFIAIDIT
ASFSTPLEEANFEFKIWNNTQEEIQKLEKEVETHPNGTKVYKTTGNLYIV
QLQIDLLTHAKLYNLNATASPTSTPTSPRRRRTSSVSGGNGDNDSTKEES
VAAQVQFAYDNLKRSINVLGELRAWCEQYRASLWYELSKINPTSVMSAYGRP
VSARKLIGVVSVDCISVDKSVFHKMKVKGEDLCYTPVPVGFKNGLSELPAGQ
LGPNEIVLSTQVEVCOHSCHEVFOAGNQYKYDYIYSTLNLDIPIHLTHMITLN
LSLVNDFKVELYKTEKELSNVEDIETMERRYNTQNLGRLKDDSDIDGRD
SFLOTGLDLMQDLGIGVNVNVAAGVSLFSGVSGVIFSNFPGMLLVLIAG
VVYVLPMTSRSSIIISAPRLMPGVERAAQEPGAPVSEDIQRLNLMGHQFOQOR
AEERAREEVAGKRTLEFVIRDSATSVLRRRRGGGYQLQRDGDSDDEYELRLQ
DGGIDDDVDEAGTADTGV"
34855 .34860

polyA_signal

/note="orf 07 and 08"

CDS

35089 .38115

/note="orf 09"

/codon_start=1

/product="DNA polymerase replicative subunit"

/protein_id="AAC13796.1"

/db_xref="GI:695181"

/translation="MSFYNPLYVKRTFLKKAAPSRTKEYTRIIPKCFKTPGAAGVVP
HTSLDPCVFGDKETPLIXGDSRLWSAGRGPGTGAGQGHFTVALTFHYVDIIE
TVYGQDRDHVPFQDIIIPSGTVLKLGRTSDDRSCVNVFROELFYVVRPEGLK
LDPLIOQCSRENFSQGRYREKTSKRYLCVVEAREVYRVFASOGFVDLILAGL
TAAGCEVFETNVDAARFIIDNGFTFGWYSCAAAPROGGAABDSWTELYDCAAGD
LEFHAGADMPCYNVLSFDEICLGENGFENASDEDMILQISCVWKAGSGEAPRSVL
NLNGTCEEIEGVVYVQPSLDLILYLFTHMRDADVEFTGINISNDFPYVIDRATQ
VYNLNLKEFTRVRSSIFEVHKFNSAGFMKAVSKVAVGVVPMYQVCRDLKLS
NYKLDTVAGECVGAKKEDVSIKEIPLHFGQGGARGLCYVKDSALVLDLLRYMT
HVEISIRAKIPTRRVLTGQOIRVFSCLLDVAGREGYILFVDRHADAEYQCATV
IDSPGFYNTPLVVDVFPYPTIQAHLNLCYSTMIPGDRCLHPLHLPGLGPDYETFELA
SGPVHEVKKHAKLALTNLWAKRAIRRELATVSDAVRTILDKQOILAIVTCN
AVYFTGVASGILFCLKIAETVTFQGRMLNKKRYIEGVTPGLADILGRVECAPD
ASFKEYIGDTSLSLTHCRGYRPEQVTGFCDELAAMTRILFVDPKLEAKETFKLIL

Query Match	13.8%	Score 32.4;	DB 16;	Length 184427;
Best Local Similarity	48.9%;	Pred. No. 13;		
Matches	87;	Conservative	0;	Mismatches 91; Indels 0; Gaps 0;
Qy	12	GCTGCCCTTCTCTCTTGTACCTCTTGGAGCTCACATGAAGAGCGCGGGTAT	71	
Db	86876	GGCTCCCCCGATCTGATCAACATCTCTGTCATCTCTCTGTCACCCCTGCGCCCT	86817	
Qy	72	GACTTTGCACTGAAGTGAAGAGTCTTTTCTGACAAATCTCTCTATGATGTCAGCTT	131	
Db	86816	GGTTCGGCTCTCGGGATGACCTGGTCTCCCCACCAAGCCCTCTACGAGCCCATCCA	86757	
Qy	132	CCTGGAATTCCTTGAAGCTCTGCTCTCTCTCCATCTCCCTTACAGGACACGCTC	189	
Db	86756	GCTGAAGAACATCAAGCAGCGGGTGAAGCTGCAGTCCCGAGCGGCGCCAGCATC	86699	

Result No.	Score	Query			ID	Description
		Match	Length	DB		
1	234	100.0	439	1	V84366	Human stomach carcinoma
2	30	12.8	7824	1	Q74052	Human interleukin-13
C 3	30	12.8	14690	1	X22303	Human IL-1ra BAC clone
C 4	29.8	12.7	291	1	V86365	EST clone AR34, New York
C 5	29.4	12.6	6175	1	T34291	Coding sequence for human
C 6	29	12.4	3580	1	V81742	Human PTP014 encoding
C 7	28	12.0	5018	1	N70646	Clone 2356p102 encoding
C 8	28	12.0	5018	1	N50150	Sequence of the cDNA
C 9	28	12.0	5962	1	Q70729	Sequence of the cDNA
C 10	28	12.0	5962	1	T42215	TATTA-binding protein
C 11	28	12.0	5962	1	T79596	Human TATTA-binding
C 12	27.8	11.9	2782	1	V71227	TATTA-binding protein
C 13	27.6	11.8	2635	1	T95763	CDNA encoding a mu
C 14	27.6	11.8	1853	1	V06751	Arabidopsis SCARCB1
C 15	27.6	11.8	1293	1	V61832	Dentatorubral pallidum
C 16	27.6	11.8	1697	1	V61833	Coding sequence for
C 17	27.4	11.7	2125	1	T87016	Coding sequence for
C 18	27.4	11.7	10023	1	V52731	Human FC-gamma receptor
C 19	27.4	11.7	984	1	V52663	Human hepatocyte nuclear
C 20	27.2	11.6	2121	1	T12262	Human native hepatoma
C 21	27	11.5	5287	1	Q25440	Mouse TRAF2 CDNA.
C 22	27	11.5	2459	1	Q95437	E. coli DNA polymerase
C 23	27	11.5	2662	1	T08977	Actual polymerase
C 24	27	11.5	1692	1	T60974	Actual shortened B
C 25	27	11.5	1785	1	T66986	NLRK2 CDNA 3' region
C 26	26.8	11.5	2259	1	Q05106	Human brain-specific
C 27	26.8	11.5	2306	1	Q05873	Sequence encoding
C 28	26.8	11.5	3979	1	Q06312	Sequence encoding
C 29	26.8	11.5	1083	1	T12910	Sequence of plasmid
C 30	26.8	11.5	1128	1	V28846	VNTR locus D6S22 f
C 31	26.8	11.5	11624	1	V52850	Mouse coxsackievirus
C 32	26.8	11.4	999	1	Q12870	Human eyal gene co
C 33	26.6	11.4	809	1	V37133	Promoter sequence
C 34	26.4	11.3	2483	1	Q14425	DNA sequence used
C 35	26.4	11.3	37913	1	Q51128	IK negative trkB g
C 36	26.4	11.3	37895	1	T34137	Eicosapentaenoic a
C 37	26.4	11.3	37895	1	T795671	Gene for biosynthe
C 38	26.4	11.3	37895	1	V00503	BRCA2 cancer suscep
C 39	26.4	11.3	808	1	V28688	S. putrefaciens cu
C 40	26.4	11.3	808	1	V28688	Rifining banana cu

1	ATGGGCTGTGGCTGCCCCCTTGTCCTCTTGACCTCCTTTGGCAGCTCACATGGACA	60
QY		
75	ATGGGCTGTGGCTGCCCCCTTGTCCTCTTGACCTCCTTTGGCAGCTCACATGGACA	134
Db		
61	GGCCGGGTATGACTTTGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCTAT	120
QY		
135	GGCCGGGTATGACTTTGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCTAT	194
Db		
121	GAGTCCACGCTCTCTGGAATTGGTTGAAAGCTGCTGCTCCTCCATCCTCCCTCAGG	180
QY		
195	GAGTCCACGCTCTCTGGAATTGGTTGAAAGCTGCTGCTCCTCCATCCTCCCTCAGG	254
Db		
181	ACACGGGTACCCCTCCACCATGCAAGATCTCAACACCATGTTGTCTGCAACACA	234
QY		

```
Db 255 ACCAGGTCACCTCCACCATGACATCTACACACCTGTTGTTCTGCACACA 308
|||||
RESULT 2
ID Q74052 standard; DNA: 7824 BP.
AC Q74052;
DE 29-JAN-1996 (first entry)
DE Human interleukin-1-beta
KW Interleukin 1 beta; primer; mRNA; specificity; pharmaceutical; ss.
OS Homo sapiens. Location/Qualifiers
FH Key misc_difference 1331
FT /*tag= a
FT /note= "N is undefined"
FT J07123984-A.
PD 16-MAY-1995.
PE 05-NOV-1993: 275852.
PR 05-NOV-1993: JP-275852.
PA (HITB ) HITACHI CHEM CO LTD.
DR WPI: 95-211627/28.
PT A primer for the detection and the determ. of a specific messenger
PT RNA - can detect and determine specific mRNA(s) with high
PT reliability
PS Example 18; Page 17-20; 35pp; Japanese.
CC Q74052 is interleukin-1-beta cDNA and Q74019-21 are primers used for
CC the amplification of this cDNA. They are used specifically for the
CC detection and isolation of this sequence. The primers have the
CC advantage of high sensitivity and reliability and are useful in the
CC pharmaceutical industry.
CC Sequence 7824 BP; 2099 A; 1905 C; 1624 G; 2195 T;
SQ

Query Match 12.8%; Score 30; DB 1; Length 7824;
Best Local Similarity 55.9%; Pred. No. 5.6;
Matches 57; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 73 ACTTTGCACTGAGCTGAGGAGCTCTTTCTGACAAATCTCTCTATGAGTCCAGCTTC 132
DB 4343 ACTCTACAGCTGGAGGTAGTGAATGCTATGGAATGAGCCCTTCTCAGCTCTGCTAC 4284
QY 133 CTGGAATTGTTGAAAGCTCTGCTCTCTCTCTCCATCTCCCT 174
DB 4283 CACTTATTTCCAGACAACACCTTCTCCGCGCCCATCCCT 4242

RESULT 4
V86365/c
ID V86365 standard; cDNA: 291 BP.
AC V86365;
DE 27-APR-1999 (first entry)
DE EST clone AR34.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PS WO9845435-A2.
PD 15-OCT-1998.
PR 10-APR-1998; US-835913.
PR 10-APR-1997; US-835913.
PA (GEM ) GENETICS INST INC.
PI Acostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 99-070076/06.
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 210; 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC therapy. The EST sequences are also stated to be useful for gene
SQ
Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 12.7%; Score 29.8; DB 1; Length 291;
Best Local Similarity 70.2%; Pred. No. 1.9;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 69 TATGACTTTCGAGCTGAGGAGCTCTTTCTGACAAATCTCTCTATGAGTTC 125
DB 257 TTGATTTTGTACAGAACGAGTGAATTTTCTTACACATTTCTCTTAAATC 201

member of the cytokine superfamily that is expected to inhibit
inflammation by binding to the interleukin-1 receptor (IL-1R). It may
also bind to a new receptor so could regulate other cellular processes
associated with acute or chronic inflammation, e.g. asthma, chronic
myelogenous leukaemia, rheumatoid arthritis, psoriasis, inflammatory
bowel disease. It may also induce or suppress interleukins, cytokines and
growth factors. Modulators of this protein are used to treat or prevent
conditions associated with abnormal levels of inflammation, or activity
of IL-1 or its receptor complex.
Sequence 14690 BP; 3810 A; 3245 C; 3299 G; 4327 T;
SQ
```

```

RESULT 5
T34291/c
ID T34291 standard; cDNA to mRNA; 6175 BP.
AC T34291;
DT 23-MAR-1997 (first entry)
DE Coding sequence for smooth muscle myosin heavy chain SM1.
KW Smooth muscle; myosin heavy chain; SM1 isoform; rabbit; arteriosclerosis;
KW gene therapy; mouse; SM2 isoform; retrovirus; adenovirus; restenosis;
KW associated adenovirus; coronary artery catheterisation; sclerotic artery;
KW ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT 105..6023
FT /*tag= a
FT /product= SM1
PN WO9623069-A1.
PD 01-AUG-1996.
PF 25-JAN-1995; J00134.
PR (OSAP) OSAKA PREFECTURAL GOVERNMENT.
PA (VSS-) VESSEL RES LAB CO LTD.
PI Arakawa E, Hasegawa K, Ishiyama H, Matsuda Y, Oda S;
PI Sugawara M, Takahashi K;
DR WPI; 96-362693/36.
DR P-PSDB; W00024.
PT DNA encoding smooth muscle myosin heavy chain SM1 isoform - used in
PT a vector for gene therapy for reduction of re-stenosis following
PT coronary artery catheterisation
PS Claim 5; Page 14-27; 42pp; Japanese.
CC This sequence represents the coding sequence for the smooth muscle myosin
CC heavy chain SM1 isoform protein. This SM1 coding sequence was isolated
CC from a mouse embryo cDNA library using probes based on the rabbit smooth
CC muscle myosin heavy chain SM2 isoform. This sequence is included in the
CC gene therapy vector of the invention. The vector of the invention
CC consists of this sequence inserted into a retrovirus, adenovirus,
CC associated adenovirus or animal-expression plasmid vector (such as pCMX2
CC or pAGE208). The vector can be used in the gene therapy treatment of
CC arteriosclerosis, particularly for the reduction of restenosis occurring
CC following coronary artery catheterisation for widening of sclerotic
CC arteries.
SQ Sequence 6175 BP; 1846 A; 1455 C; 1856 G; 1018 T;

Query Match 12.6%; Score 29.4; DB 1; Length 6175;
Best Local Similarity 54.1%; Pred. No. 7.9;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 124 TCCAGCTCTCGAATTCGTTGAAGAGCTGCTGCTCCCTCCCTCCATCTCCCTTCAGGGACC 183
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
302 TCCATTTTCCACCACTCCACGACCACTCATCGCCCTTCTCTCTCTGATGCTGCTGC 243
QY 184 AGCGTCACTCCACCACTGCAAGATCTCAACACCACTGTTGTCGCAACACA 234
DB || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
242 TTCGAAGCCCTGCTTCTGTAAGGAGGCCACACCACTGCTTCTGGTACCCA 192

RESULT 6
V81742/c
ID V81742 standard; cDNA; 3580 BP.
AC V81742;
DT 10-MAR-1999 (first entry)
DE Human PTP04 encoding CDNA.
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 53..2476
FT /*tag= a
FT /product= a
PN WO9849317-A2.
PD 05-NOV-1998.

```

```

PF 27-APR-1998; U08439.
PR 23-OCT-1997; US-063595.
PR 28-APR-1997; US-044428.
PR 20-MAY-1997; US-047222.
PR 11-JUN-1997; US-049477.
PR 11-JUN-1997; US-049756.
PR 18-JUN-1997; US-049914.
PA (SUGEN-) SUGEN INC.
PI App H, Clary D, Courtneidge SA, Rui TH, Jallal B,
PI Markby D, Ornst S, Peles E, Plowman GD;
DR WPI; 99-009434/01.
DR P-PSDB; W89247.
PT New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
PS Claim 2; Page 145-146; 193pp; English.
CC The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes human PTP04. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
SQ Sequence 3580 BP; 1213 A; 666 C; 618 G; 1083 T;

Query Match 12.4%; Score 29; DB 1; Length 3580;
Best Local Similarity 48.0%; Pred. No. 8.6;
Matches 83; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 23 TCCTCTCTTGACCTCTCTGCGAGCTCACATGGAACAGGGCGGTATGACTTTGCAAC 82
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
642 TCATGCTGCGCAATTCCTGTAATGAACAGGTAGATGCTGAGTTCTACTATTGAAC 583
QY 83 TGAAGTGAAGAGAGTCTTTCTGACAAATTCCTCTATGATCCAGCTCTCTGGAATTGC 142
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
582 TTAACCTTTAGAGTCCGATATATATACAGATTTCTTTTTCAGCTTCACAGATACA 523
QY 143 TTGAAAGCTCTGCTCTCTCTCCATCTCCCTTCAGGGACCGGTACACCTC 195
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
522 GAGAAAGGCCAAATTCACGCTGCATCTCTCTGCTCAGCCCACTAGGCTC 470

RESULT 7
N70646/c
ID N70646 standard; cDNA; 5018 BP.
AC N70646;
DT 27-APR-1991 (first entry)
DE Clone 23B6p102 encoding polypeptide with IgE binding factor
DE activity.
KW Antibody; immunoglobulin; ss.
FH Key Location/Qualifiers
FT 423..2753
FT /*tag= a
FT /product= a
PN JG2045600-A.
PD 27-FEB-1987.
PF 21-AUG-1985; 183810.
PR 21-AUG-1985; JP-183810.
PA (SCHE-) SCHERING BIOTECH CO.
DR WPI; 87-097765/14.
DR P-PSDB; P70417.
PT CDNA clone coding polypeptide - showing IgE bond factor activity.
PS Disclosure; Page 29; 33pp; Japanese.
CC The clone encodes a polypeptide which shows IgE binding factor

```



```

AC T95763;
DT 11-MAY-1998 (first entry)
DE Arabidopsis SCARECROW clone SRP12.
KW SCARECROW; SCR; SRP12 gene; transgenic plant; root; gravitropism;
KW crop improvement; ss.
OS Arabidopsis thaliana.
FH Key Location/Qualifiers
FT CDS 3..2429
FT /*tag= a
PN W09741152-A1.
PD 06-NOV-1997.
PE 25-APR-1997; U07022.
PR 24-APR-1997; US-842445.
PR 26-APR-1996; US-638617.
PA (UNYV) UNIV NEW YORK STATE.
PI Benfey PN, Dilaurenzio L, Helariutta Y, Malamy JB,
PI Pysh L, Wysocka-Diller J;
DR WPI; 97-549683/50.
DR P-FSDB; W38188.
PT DNA encoding Arabidopsis SCARECROW protein - useful to modify plant
PT cell division and therefore alter root development, or alter plant
PT stem or hypocotyl gravitropism
PS Claim 6; Page 130-131; 221pp; English.
CC This sequence comprises a partial cDNA clone of the Arabidopsis
CC SCARECROW (SCR) SRP12 gene, identified by homology to the
CC Arabidopsis SCR gene (see T95753). It encodes a 808-amino acid
CC polypeptide (see W38188). SCR is expressed specifically in embryo
CC root progenitor tissue and in certain root and stem tissues. It
CC regulates a specific asymmetric division, and controls gravitropic
CC response in aerial structures and root formation. Nucleic acid
CC molecules (see T95753-66) encoding SCR protein sequences (see
CC W38178-201), SCR proteins lacking 1 to 4 of motifs I to VI, or
CC SCR motif I, II, III, IV, V or VI are claimed. Transgenic plants
CC can be engineered to overexpress SCR, so that cell division is
CC increased in roots, resulting in thicker root development, while a
CC plant with an altered stem or hypocotyl gravitropism is less
CC susceptible to lodging. Plants that contain an antisense
CC molecule that suppresses the expression of endogenous SCR gene
CC product show thinner root development. A gene of interest can be
CC placed under control of a SCR promoter and expressed in a plant to
CC confer herbicide, salt, pathogen or insect resistance, or when
CC expressed in stems to increase starch, lignin or cellulose
CC biosynthesis (all claimed).
SQ Sequence 2635 BP; 696 A; 501 C; 623 G; 778 T;

Query Match 11.8%; Score 27.6; DB 1; Length 2635;
Best Local Similarity 50.8%; Pred. No. 21;
Matches 66; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 72 GACTTTGCACTGAAGCTGAGGAGTCTTTCTGACAAATTCCTCTATGAGTCCAGCTT 131
Db 2301 GATTTCTGCATCAGTTCCTTCTCAAGCGGAGCTGTCTAATCCGGCTCGGATCAGTCTCG 2242
Qy 132 CTGGAATTCCTGAAAGCTTCGCTCTCTCCATCTCCCTTCAGGAGCCAGCTCAC 191
Db 2241 CTGGCACTGCTTATATGCTCTCTCGCTCTCCACTCTCTCTGTTCTCTCAGAGCCACAA 2182
Qy 192 CTTCCACCAT 201
Db 2181 CATTCAACAT 2172

RESULT 14
V06751
ID V06751 standard; DNA; 1853 BP.
DE Dentatorubral-pallidolusian atrophy cDNA (nucleotides 238-2090).
KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; polyglutamine;
KW CAG repeat domain; neurodegenerative disorder; inhibition; DRPLA;
KW Huntington's disease; schizophrenia; psychiatric disorder;
KW dentatorubral-pallidolusian atrophy; ss.

OS Homo sapiens.
PN US5723301-A.
PD 03-MAR-1998.
PE 03-NOV-1995; 553110.
PR 03-NOV-1995; US-553110.
PA (YUDU-) UNIV DUKE.
PI Burke JK, Enghild J, Strittmatter WJ, Vance JM;
DR WPI; 98-178475/16.
PT Screening assay for inhibitors of GAPDH binding to polyglutamine -
PT for use in treating neuro-degenerative diseases or psychiatric
PT disorders
PS Example 7; Columns 21-24; 19pp; English.
CC This sequence represents dentatorubral-pallidolusian atrophy (DRPLA)
CC cDNA (nucleotides 238-2090 of Genbank accession number D38529). The
CC invention relates to a new screening assay for compounds capable of
CC inhibiting binding of glyceraldehyde-3-phosphate dehydrogenase (GAPDH)
CC to polyglutamine. The assay comprises providing an aqueous solution
CC containing a test compound, polyglutamine-containing molecules and GAPDH
CC and determining if binding of GAPDH to the polyglutamine-containing
CC molecules is reduced in the presence of the test compound. Inhibitors
CC identified by the above assay could be used to treat neurodegenerative
CC diseases or psychiatric disorders caused by a gene containing an
CC extended CAG domain, e.g. Huntington's disease or schizophrenia.
CC Sequence 1853 BP; 347 A; 774 C; 388 G; 344 T;

Query Match 11.8%; Score 27.6; DB 1; Length 1853;
Best Local Similarity 56.7%; Pred. No. 18;
Matches 51; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 112 TCCTCTATGAGTTCAGCTTCCTCGAATTCCTTGAAGAAGCTTCGCTCTCTCCATCTC 171
Db 940 TCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 999
Qy 172 CCTTCAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 201
Db 1000 CTTCCCTCCCAACGAGCTCTCTGTCCTCAAT 1029

RESULT 15
V61832/c
ID V61832 standard; cDNA; 1293 BP.
AC V61832;
DT 28-APR-1999 (first entry)
DE Coding sequence for human p53 regulated protein, P2XM.
KW Human; p53 regulated protein; P2XM; ATP receptor; Rp-2 protein;
KW thymocyte; apoptosis; tumour-suppressor gene; rhabdoid sarcoma;
KW anticancer drug; ss.
OS Homo sapiens.
PN W09842835-A1.
PD 01-OCT-1998.
PE 18-MAR-1998; J01146.
PR 26-MAR-1997; JP-093044.
PA (SAKA) OTSUKA PHARM CO LTD.
PI Nakamura Y, Iwano T;
DR WPI; 98-532006/45.
DR P-FSDB; W76434.
PT Human gene P2XM whose transcription is induced by p53 - useful, e.g.
PT for diagnostic purposes and in development of new anticancer drugs
PS Claim 2; Page 22-23; 43pp; Japanese.
CC This sequence represents the coding sequence for the P2XM protein of the
CC invention. The protein is significantly homologous to: (i) the P2X family
CC of ATP receptors, and (ii) Rp-2 protein which is expressed in thymocytes
CC during apoptosis. Transcription of the genes is specifically regulated by
CC the tumour-suppressor gene p53. The P2XM gene is specifically expressed
CC in skeletal muscle and has been localised to chromosome 22q11, an area
CC where mutation and sequence losses frequently occur in rhabdoid sarcomas.
CC The genes may be used for diagnostic purposes (e.g. by detecting changes
CC occurring in the gene in sarcomas), using probes and primers containing
CC or derived from all or part of the genes. The genes may further be used
CC in the development of new anticancer drugs.
CC Sequence 1293 BP; 271 A; 379 C; 389 G; 254 T;

```

Query Match 11.8%; Score 27.6; DB 1; Length 1293;
Best Local Similarity 58.5%; Pred. No. 16;
Matches 48; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 136 GAATTGCTTGAAGAAGCTCTGCTCTCTCTCCATCTCCTCAGGGACCGTCACCTC 195
DB 567 GAACAGTGTGAAGTTCTGGGCTGGGCGCAGCAGGGGCTCGAGGGCACACCCCACTCTC 508
QY 196 CACCATGCAAGATCTCAACCC 217
DB 507 CACGGGCGACCACTCCAGATC 486

Search completed: March 20, 2000, 09:31:54
Job time: 1077 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2000, 15:56:36 ; Search time 61.26 Seconds
(without alignments)
457.314 Million cell updates/sec

Title: US-09-092-296-7_COPY_51_284

Perfect score: 234

Sequence: 1 ATGGGGTCTGGGTGCCCTCTCTTTCACACACACA 234

Scoring table: IDENTITY_NUC

Searched: 214294 seqs, 59861208 residues

Database : Issued_Patents_NA.*

Word size : 0

Number of hits that pass the threshold : 428588

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
- 7: /cgn2_6/ptodata/2/ina/PTUS9_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30.8	13.2	7218	1	US-08-232-463-14
2	28.6	12.2	2662	5	US-08-750-357-8
3	28	12.0	5962	1	US-08-188-582-10
4	28	12.0	5962	1	US-08-646-715-10
5	27.8	11.9	2782	3	US-08-937-466-1
6	27.8	11.9	2782	4	US-09-172-528-1
7	27.6	11.8	1853	2	US-08-553-110-2
8	27.2	11.6	2121	1	US-08-331-394-3
9	27.2	11.6	2121	2	US-08-250-858-3
10	27.2	11.6	2121	2	US-08-446-915-3
11	27.2	11.6	2121	3	US-08-744-139-3
12	27.2	11.6	2121	6	PCR-US95-06639-3
13	27	11.5	2459	1	US-08-101-593-5
14	27	11.5	2459	1	US-08-485-995A-5
15	27	11.5	2459	4	US-08-485-994C-5
16	27	11.5	1960	4	US-08-766-858A-4
17	26.8	11.5	2259	7	5185254-3
18	26.6	11.4	999	2	US-08-416-478A-4
19	26.6	11.4	999	3	US-08-474-988B-4
20	26.6	11.4	999	4	US-08-394-442B-4
21	26.4	11.3	2463	1	US-08-339-578-1
22	26.4	11.3	37895	1	US-08-375-709-1
23	26.4	11.3	8268	1	US-08-375-709-10
24	26.4	11.3	37895	2	US-08-752-929-1
25	26.4	11.3	8268	2	US-08-752-929-10
26	26.2	11.2	441	1	US-08-149-091-4
27	26.2	11.2	441	4	US-08-637-761-4
28	26.2	11.2	441	6	PCR-US94-12672-4
29	26	11.1	2496	1	US-08-073-384C-2
30	26	11.1	9278	1	US-08-243-542-9
31	26	11.1	2496	1	US-08-254-359A-2
32	26	11.1	9278	1	US-08-477-407-9

33	26	11.1	2496	2	US-08-483-043-2	Sequence 2, Appl 1
34	26	11.1	9278	2	US-08-484-355-9	Sequence 9, Appl 1
35	26	11.1	2496	2	US-08-481-238-2	Sequence 2, Appl 1
36	26	11.1	1370	2	US-08-592-126-79	Sequence 79, Appl 1
37	26	11.1	2496	3	US-08-471-066B-2	Sequence 2, Appl 1
38	26	11.1	2496	3	US-08-484-956-2	Sequence 2, Appl 1
39	26	11.1	2496	3	US-08-757-653-2	Sequence 2, Appl 1
40	26	11.1	2496	3	US-08-599-491-2	Sequence 2, Appl 1
41	26	11.1	2496	4	US-08-756-386-2	Sequence 2, Appl 1
42	26	11.1	2496	4	US-08-823-516-2	Sequence 2, Appl 1
43	26	11.1	2496	5	US-08-682-853A-2	Sequence 2, Appl 1
c 44	25.8	11.0	2334	1	US-08-406-070-1	Sequence 1, Appl 1
45	25.4	10.9	1560	1	US-07-669-171-1	Sequence 1, Appl 1

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 13.2%; Score 30.8; DB 1; Length 7218;
Best Local Similarity 3.5%; Pred. No. 0.89;
Matches 8; Conservative 128; Mismatches 90; Indels 0; Gaps 0;

QY 1 ATGGGGTCTGGGTGCCCTCTCTTTCACACACACA 60

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA: US/09/172,528
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/937,466
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: OSMAN, RICHARD A.
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: T97-009
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (650) 343-4341
;; TELEFAX: (650) 343-4342
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2782 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
US-09-172-528-1

Query Match 11.9%; Score 27.8; DB 4; Length 2782;
Best Local Similarity 53.2%; Pred. No. 5.7;
Matches 59; Conservative 0; Mismatches 0; Gaps 0;

QY 56 GAACAGGGCGGTATGACTTGGCACTGAGCTGAAGAGTCTTTTCGACAAATTCCT 115
DB 2101 GAAACCTCGGAGAGAGGAGCAGAGCAGCTGTTCTTCCTTGGCCCTAAGGTGGAG 2042

QY 116 CCGTAGTCCAGCTCTCTGGAATGCTTGAAGAAGCTGCTCCCTCCCTCC 166
DB 2041 ATTCTGAGTCTCTGCTTGAAGTATGATGAGCAGCAGTGCCTCCCTCC 1991

RESULT 7
US-08-553-110-2
; Sequence 2, Application US/08553110
; Patent No. 5723301
; GENERAL INFORMATION:
; APPLICANT: Burke, James R.
; APPLICANT: Vance, Jeffrey M.
; APPLICANT: Enghild, Jan
; APPLICANT: Stittmatter, Warren J.
; TITLE OF INVENTION: Therapeutics for Diseases Associated
; WITH CAG Trinucleotide Repeat Expansion
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley
; STREET: P.O. Drawer 34009
; CITY: Charlotte
; STATE: No. 5723301th Carolina
; COUNTRY: United States of America
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,110
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-117
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200

;; TELEFAX: 919-881-3175
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1853 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-553-110-2

Query Match 11.8%; Score 27.6; DB 2; Length 1853;
Best Local Similarity 56.7%; Pred. No. 5.5;
Matches 51; Conservative 0; Mismatches 0; Gaps 0;

QY 112 TCCTCTATGAGTCCAGCTTCTCGGAATTGCTGAAAGCTCTGCTCCCTCCCTCCATCTC 171
DB 940 TCCTCTCTGCTCCCTCCCTTCCAGCATTCCCGAGCATTCGCCAGCTACCCGCCACTCTTTC 999

QY 172 CCTTCAGGAGCAGCTCACCTCCACCAT 201
DB 1000 CTCCCCCAACAGCCTCTCTGTCTCCAAT 1029

RESULT 8
US-08-331-394-3/c
; Sequence 3, Application US/08331394
; Patent No. 5670319
; GENERAL INFORMATION:
; APPLICANT: Goeddel, David V.
; APPLICANT: Rothe, Mike
; TITLE OF INVENTION: Tumor Necrosis Factor
; TITLE OF INVENTION: Receptor-Associated Factors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,394
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/250858
; FILING DATE: 27-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 897p1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2121 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-331-394-3

Query Match 11.6%; Score 27.2; DB 1; Length 2121;
Best Local Similarity 53.8%; Pred. No. 8;
Matches 56; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

RESULT 12
PCT-US95-06639-3/c
Sequence 3, Application PC/US9506639
GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
TITLE OF INVENTION: Tumor Necrosis Factor Receptor-Associated Factors
NUMBER OF SEQUENCES: 66
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06639
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/250858
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/331394
FILING DATE: 28-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Dregler, Ginger R.
REGISTRATION NUMBER: 33,055

```

RESULT 13
US-08-101-593-5
; Sequence 5, Application US/08101593
; Patent No. 5547859
; GENERAL INFORMATION:
; APPLICANT: Goodman, Myron F.
; APPLICANT: Reha-Krantz, Linda J.
; TITLE OF INVENTION: NEW DNA SEQUENCING ENZYMES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Robbins, Berliner & Carspm
; STREET: 201 No. 5547859th Figueroa Street, Fifth Floor
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90012-2628
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/101,593
; FILING DATE: 19930802
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Spitals, John P.
; REGISTRATION NUMBER: 29,215
; REFERENCE/DOCKET NUMBER: 1920-305
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 977-1001
; TELEFAX: (213) 977-1003
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2459 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomlc)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..2456
; US-08-101-593-5

```

Query Match: 11.5%; Score 27; DB 1; Length 2459;

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 19, 2000, 08:46:58 ; Search time 538.4 Seconds
(without alignments)
1640.983 Million cell updates/sec
Title: US-09-092-296-7_COPY_51_284
Perfect score: 234
Sequence: 1 ATGGGGTCTGGGTCGCCCT.....ACCATGTTGTCGCAACACA 234
Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*

- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss13:*
- 103: gb_gss14:*
- 104: gb_gss15:*
- 105: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	175	74.8	404	61	AI857998
C 2	93.2	39.8	328	42	AI136523
C 3	75.8	32.4	552	82	AQ718761
C 4	36.4	15.6	331	31	AA282040
C 5	36.4	15.6	494	34	AA534569
C 6	36.4	15.6	424	38	AA741185
C 7	36.4	15.6	402	40	AA936790

c 8	36.4	15.6	529	41	A1018775
c 9	34.8	14.9	451	61	A1829206
c 10	34	14.5	190	41	A1010074
c 11	33.4	14.3	545	23	T41524
c 12	33.2	14.2	292	40	A8991822
c 13	33.2	14.2	597	63	A8991822
c 14	32.6	13.9	522	82	AQ1718755
c 15	32.4	13.8	490	51	A1757641
c 16	32.2	13.8	875	82	AQ740265
c 17	32	13.7	573	23	T41535
c 18	31.8	13.6	547	23	T41523
c 19	31.8	13.6	267	50	AU062745
c 20	31.8	13.6	613	63	A1981365
c 21	31.8	13.6	568	100	AQ290553
c 22	31.6	13.5	502	102	AQ373909
c 23	31.4	13.4	505	20	M89011
c 24	31.4	13.4	433	23	R90433
c 25	31.4	13.4	523	23	T44883
c 26	31.4	13.4	548	23	T76643
c 27	31.4	13.4	529	37	H76456
c 28	31.4	13.4	497	37	T44613
c 29	31.2	13.3	615	59	A1765674
c 30	31.2	13.3	617	64	AL042852
c 31	31.2	13.3	637	64	AW025169
c 32	31	13.2	398	20	T52782
c 33	31	13.2	307	20	Z45258
c 34	31	13.2	536	23	T41525
c 35	31	13.2	466	60	A1803834
c 36	30.8	13.2	459	41	A1442028
c 37	30.8	13.2	411	62	AU076246
c 38	30.6	13.1	313	20	T36119
c 39	30.6	13.1	368	21	R15181
c 40	30.6	13.1	467	21	T59783
c 41	30.6	13.1	441	22	R34436
c 42	30.6	13.1	437	23	AA040444
c 43	30.6	13.1	257	33	AA443415
c 44	30.6	13.1	363	43	A1219495
c 45	30.6	13.1	471	50	A1692905

ALIGNMENTS

RESULT 1	A1857998/c	404 bp	mRNA	EST	26-AUG-1999
LOCUS	wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2408041 3',				
DEFINITION	mRNA sequence.				
ACCESSION	A1857998				
VERSION	A1857998.1	GI:5511614			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 404)				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
COMMENT	Unpublished (1997) On Jun 5, 1998 this sequence version replaced gi:3189004. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html				

Seq primer: -40UP from Gibco	
High quality sequence stop: 395.	
Location/Qualifiers	
1. 404	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:2408041"	
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"	
/dev_stage="adult"	
/lab_host="DH10B (phage-resistant)"	
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dr) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	97 a 105 c 117 g 84 t 1 others
ORIGIN	
Query Match	74.8%; Score 175; DB 61; Length 404;
Best Local Similarity	99.4%; Pred. No. 6.6e-41;
Matches 175; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	59 CAGGCGCGGTATGACTTCAACTGAGGAGTCTTTCTGCAAAATTCCTCT 118
Db	311 CAGGCGCGGTATGACTTCAACTGAGGAGTCTTTCTGCAAAATTCCTCT 252
QY	119 ATGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178
Db	251 ATGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 192
QY	179 GGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 234
Db	191 GGACGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 136

RESULT 2	A1136523/c	328 bp	mRNA	EST	11-FEB-1999
LOCUS	UI-R-C2p-bq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus CDNA clone				
DEFINITION	UI-R-C2p-bq-e-02-0-UI 3', mRNA sequence.				
ACCESSION	A1136523				
VERSION	A1136523.1	GI:3637300			
KEYWORDS	EST.				
SOURCE	Norway rat.				
ORGANISM	Rattus norvegicus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.				
AUTHORS	1 (bases 1 to 328)				
TITLE	Bonaldo, M.F., Lennon, G. and Soares, M.B. Normalization and subtraction: two approaches to facilitate gene discovery				
JOURNAL	Genome Res. 6 (9), 791-806 (1996)				
MEDLINE	97044477				
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2150222. Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Tel: 319 335 8250 Fax: 319 335 9365 Email: msoares@uiowa.edu The sequence tag present in the cDNA between the NotI site and the oligo-dr track used to identify it as a clone from the normalized adult lung library. cDNA Library Preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research				

```
Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1. 328
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-nq-e-02-0-UI"
/clone.lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site1: Not I; Site2: Eco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dr track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)."

BASE COUNT      62 a  77 c  98 g  91 t
ORIGIN

Query Match      39.8%; Score 93.2; DB 42; Length 328;
Best Local Similarity 68.8%; Pred. No. 3.6e-17;
Matches 128; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY  44 GCAGCTCATGGAACAGCGGGGTATGACTTTGCAACTGAAGCTGAAGAGCTCTTTC 103
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  327 GCAGCTCACTGAGAGGCCCAACAAAGGTGACTTTGCAGTAACATGACCGAGCATCTC 268
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  104 TGCAAAATTCCTCTATGAGTCCAGCTCTCTGGAAATGCTTGAAAGCTCTGCTCTCC 163
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  267 AGGCCAAGACCTCCCAAGACTCCGGCTTTCTGGACATGCTCCAAAAGATCTGCTCTCC 208
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  164 TCCATCTCCCTTACGAGGACGAGCTCAGCCCTCCACCATCAAGATCTCAACACCATGTTG 223
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  207 TCCACTCTCACCAGGACCATGTCACCCCTTCATATAAGGGCCACCAACCACTTA 148
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  224 TCTGCA 229
      |||||
Db  147 CTTGCA 142

RESULT  3
LOCUS   AQ718761/c
DEFINITION HS_5511_B2_F09_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-1087 Col-18 Row-L, genomic survey sequence.
ACCESSION AQ718761
VERSION   AQ718761.1 GI:5468077
KEYWORDS  GSS.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 552)
AUTHORS  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
MEDLINE
COMMENT   Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 1087 row: L column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.
Location/Qualifiers
1. 552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1087 Col-18 Row-L"
/clone.lib="RPCI-11 Human Male BAC Library"
/sex="male"
/notes="Vector: pBACE3.6; Genomic sequence of BAC ends"

BASE COUNT      141 a  153 c  124 g  117 t  17 others
ORIGIN

Query Match      32.4%; Score 75.8; DB 82; Length 552;
Best Local Similarity 86.5%; Pred. No. 5e-12;
Matches 83; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY  57 ACAGGCGCGGTATGACTTGCACCTGAAGCTGAAGAGTCTTTCTGACAAATTCCTC 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  474 ACATGGGAGGTATGACTTTGCANCCTGAAGCTGAAGAGTCTATTATGACAAATTCCTC 415
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY  117 CTATGAGTCCAGCTCTCTGGAATTCCTTGAAGCT 152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db  414 TTATGAGTCCAGCTTCTGGAATTCCTTGAAGGT 379
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT  4
LOCUS   AA282040/c
DEFINITION z89a03.s1 NC1_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:704620 3',
mRNA sequence.
ACCESSION AA282040
VERSION   AA282040.1 GI:1924872
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 331)
AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL   On Sep 12, 1996 this sequence version replaced gi:1402231.
COMMENT   Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 891 Std Error: 0.00
Seq primer: -41m13 fwd. ET from Amersham
```

High quality sequence stop: 242.

FEATURES

source

1. .331
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="IMAGE:925940"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - oligo(dT) primer [5'-TGTACCAATCTGAAGTGGGCGCGCTCATTTTTTTTTTTTTTTT-3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 61 c 51 g 122 t

ORIGIN

Query Match 15.6%; Score 36.4; DB 31; Length 331;
Best Local Similarity 56.8%; Pred. No. 1.2;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 87 GCTGAAGGAGTCTTTCTGCAAAATTCCTCTATGATGCTCCAGTCCCTCGGAATTGCTTGA 146
|||||
Db 318 GATGATGACTGTACTGTCTATATAATCCACTAAATCCAGTACCAGGAACCTGCCTGG 259
|||||

Qy 147 AAAGCTCTGCTCTCCATCTCCCTTCAGGACCGAGCGTCCACCTCCACCATGCA 204
|||||
Db 258 AACTGTGGCCATGTCATTTTTTTTCTTAAAGACCGAGTGTGATAGTAGGCCATGCA 201
|||||

RESULT 5

AA534569/c

LOCUS n177c11.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:925940 3',
DEFINITION mRNA sequence.

ACCESSION AA534569
VERSION AA534569.1 GI:2278822

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 494)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On May 9, 1995 this sequence version replaced gi:802984.
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 370.
Location/Qualifiers

FEATURES

source

source

1. .494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from 12 pooled bulk tumor samples and primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization."

BASE COUNT 139 a 108 c 94 g 153 t

ORIGIN

Query Match 15.6%; Score 36.4; DB 34; Length 494;
Best Local Similarity 56.8%; Pred. No. 1.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 87 GCTGAAGGAGTCTTTCTGCAAAATTCCTCTATGATGCTCCAGTCCCTCGGAATTGCTTGA 146
|||||
Db 296 GATGATGACTGTACTGTCTATATAATCCACTAAATCCAGTACCAGGAACCTGCCTGG 237
|||||

Qy 147 AAAGCTCTGCTCTCCATCTCCCTTCAGGACCGAGCGTCCACCTCCACCATGCA 204
|||||
Db 236 AACTGTGGCCATGTCATTTTTTTTCTTAAAGACCGAGTGTGATAGTAGGCCATGCA 179
|||||

RESULT 6

AA741185/c

LOCUS ob30a02.s1 NCI_CGAP_Kid5 Homo sapiens cDNA clone IMAGE:1325162 3',
DEFINITION mRNA sequence.

ACCESSION AA741185
VERSION AA741185.1 GI:2779777

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 424)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Jan 5, 1998 this sequence version replaced gi:2581148.
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
unknown library type
Insert Length: 742 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 387.

Location/Qualifiers
1. .424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' AACTGAGAAATCGCGCGCAATATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library

FEATURES

source

BASE COUNT 121 a 90 c 77 q 136 t
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.

Query Match 15.6%; Score 36.4; DB 38; Length 424;
Best Local Similarity 56.8%; Pred. No. 1.3;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy	87	GCTGAGGAGTC	TTTCTGACAAAT	TTCCTCCTATGATGCCAGCTTCTCTGGAAATGCTGA	146
Db	303	GATGATGCTACTGTCATATATAAATCCACTAAATCCAGCTACCGAGGAATCGCTGG			

Oy	147	AAAGCTCTGCTCCTCCTCTCCATCTCCCTTCAAGGACCAAGCTCACCTCCACCATGCA	204
Db	243	AACTGGGCCATGCATTTTTTTTTTCTTTAAAGACCAAGTGTGATAGTAGGCCATGCA	186

RESULT	7
AA336790/c	
LOCUS	
DEFINITION	AA936790 402 bp mRNA EST 19-MAY-1998 oh89a02.s1 NCI_CGAP_Co8 Homo sapiens CDNA clone IMAGE1474154 3', mRNA sequence.

ACCESSION AA936790
VERSION AA936790.1 GI:3094824
KEYWORDS EST.
SOURCE human.

ORGANISM HOMO sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 403)

AUTHORS (Kobayashi T, Ohtsuka Y, Nakamura K, et al.)
TITLE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>. National Cancer Institute, Cancer Genome Anatomy Project (CGAP). Tumor Gene Index
JOURNAL Unpublished (1997)

On Jan 17, 1998 this sequence version replaced gi:2045507.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@NIH.gov

email: robert_strausberger@nigms.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Assayed by: M. Bento Soares, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bic.llnl.gov/ncic/

Insert Length: 676 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
high quality.

```
FEATURES
source
      ngram quality sequence stop: 386.
      Location/Qualifiers
        1..402
          /organism="Homo sapiens"
            .33t
              c"
```

```

/od_xref="taxon:9606"
/clone="IMAGE:1474154"
/clone_lib="NCI_CGAP_Co8"
/tissue_type="adenocarcinoma"

```

```

/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -

```

oligo(dT) primer. Double-stranded cDNA was ligated to RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed

BASE COUNT
ORIGIN

118 a	85 c	74 g	125 t
-------	------	------	-------

Bento Soares and M. Fatima Bonaldo.

Query Match	15.6%	Score 36.4	DB 40	Length 402	.
-------------	-------	------------	-------	------------	---

Best Local Similarity 56.8%; Pred. No. 1.2;
Matches 67; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

b

```
87 GCTGAAGGAGTCTTTTCGCACAAATCCCTCCTATGAGTCCTGGATTGCTTGA 146  
- - - - -  
296 GATGATGTACTACTCTCATATATAATCCACTAATCCAGCTACCAGGAAC TGCTGG 237  
- - - - -
```

147 AAAGCTGTGCCTCCTCTCCATCTCGCTTCAGGGACGAGCGTCAACCTCCACCATGCA 204
|| | | | | | | | | | | | | | | | | | |
236 AACTGTGCCCATGCATTTTTTTTCTTTAAAGACGAGTGATAGTAGGCCATGCA 179

RESULT 8
I018775/c

LOCUS AI018775 529 bp mRNA EST 27-AUG-1998
 ov32e12.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1639054
 3', mRNA sequence.
 ACCESSION AI018775

VERSION	AI018775.1	GI:3232573
SYNOPSIS		
SOURCE	human.	
ORGANISM	Homo sapiens	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 529)
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 14, 1998 this sequence version replaced 241.1707552

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares
Ph. D. M. Bento Soares

cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone Distribution: NCI-CGAP clone distribution information

Insert Length: 1493 Std Error: 0.00

Seq primer: -40m3 fwd. ET from Amersham
High quality sequence stop: 451.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1639054"
/clone_lib="Source: test1"

```

```

/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified

```

polymerase; site 1: Not I; site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5,

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library

SE COUNT	146 a	114 c	104 g	165 t
UGIN				

went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo. "

Query Match	15.6%;	Score 36.4;	DB 41;	Length 529;
Best Local Similarity	56.8%;	Pred. No. 1.4;		

[illegible]

Db 299 GATGATGCTACTGCTCATATATAATCCACTTAATCCAGTACCAGGACCTGCTGG 240

Qy 147 AAAGTCGCTCCTCTCTCCATCTCCCTTCAGGACACAGCTCACCCTCCACCATGCA 204

Db 239 AACTGTGGCCATGCAATTTTTTTTTTTCTTAAGACACAGCTGTGATAGTAGGCCATGCA 182

RESULT 9

AI829206/c 451 bp mRNA EST 26-AUG-1999

LOCUS wk76f01.x1 NCI_CGAP_Panl Homo sapiens cDNA clone IMAGE:2421337 3', mRNA sequence.

DEFINITION AI829206

ACCESSION AI829206

VERSION 1 GI:5449877

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3189627.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Life Technologies catalog #: 11548-013

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/obrp/image/image.html

Seq primer: -40UP from Gibco

High quality sequence stop: 426.

FEATURES

source

1..451

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2421337"

/clone_lib="NCI_CGAP_Panl"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 1.72 kb. Life Technologies catalog #:

11548-013

BASE COUNT 127 a 97 c 85 g 142 t

ORIGIN

Query Match 14.9%; Score 34.8; DB 61; Length 451;

Best Local Similarity 55.9%; Pred. No. 3.8;

Matches 66; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 87 GCTGAAGAGCTCTTTCTGCAAAATCCCTATGAGTCAGCTCTCTGGAATGCTTGA 146

Db 297 GATGATGCTACTGTCTATATAATCCACTTAATCCAGTACCAGGAACTGCCGCG 238

Qy 147 AAAGTCCTGCTCCTCCATCTCCCTTCAGGACACAGCTCACCCTCCACCATGCA 204

Db 237 ACCTGTGCCATGCAATTTTTTTTTTCTTAAGACACAGCTGTGATAGTAGGCCATGCA 180

RESULT 10

AI010074/c 190 bp mRNA EST 15-JUN-1998

LOCUS EST204525 Normalized rat lung, Bento Soares Rattus sp. cDNA clone

DEFINITION RLUB38 3' end, mRNA sequence.

ACCESSION AI010074

VERSION AI010074.1 GI:3223906

KEYWORDS EST.

SOURCE Rattus sp.

ORGANISM Rattus sp.

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

AUTHORS 1 (bases 1 to 190)

Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.

Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat Gene Index

JOURNAL Unpublished (1998)

COMMENT On Jan 19, 1998 this sequence version replaced gi:2150555.

Contact: Lee, NH

ATCC

The Institute for Genomic Research

9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529

Fax: (301)-838-0208

Email: nhlee@tigr.org

Seq primer: M13-21.

FEATURES

Location/Qualifiers

1..190

Source

/organism="Rattus sp."

/db_xref="taxon:10118"

/clone="RLUB38"

/clone_lib="Normalized rat lung, Bento Soares"

/note="Organ: lung; Vector: p773pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 39 a 51 c 59 g 41 t

ORIGIN

Query Match 14.5%; Score 34; DB 41; Length 190;

Best Local Similarity 69.7%; Pred. No. 4.7;

Matches 46; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

Qy 164 TCCATCTCCCTTCAGGACGACGTCACCTCCACCATCCAGATCTCACCACATGTTG 223

Db 190 TCCACCTCTCACCAGGACCAATGTACCCCTTCATCAATAGGGCCACACACCTTTA 131

Qy 224 TCTGCA 229

Db 130 CTTGCA 125

RESULT 11

T41524

LOCUS T41524 545 bp mRNA EST 07-AUG-1995

DEFINITION 10105 Lambda-PRL2 Arabidopsis thaliana cDNA clone 100A7T7, mRNA sequence.

ACCESSION T41524

VERSION T41524.1 GI:931185

KEYWORDS EST.

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

AUTHORS 1 (bases 1 to 545)

Newman, T., deBuijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E. and Somerville, C.

Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones

Plant Physiol. 106, 1241-1255 (1994)

95148729

JOURNAL

MEDLINE

COMMENT On Jan 24, 1995 this sequence version replaced gi:634112.

Contact: Thomas Newman

MSU-DOE Plant Research Laboratory

Michigan State University

MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI

Tel: 517-353-0854

Wed Mar 22 12:07:56 2000

Db 128 TGCCTGCTTTGGACATCTTCCTTCAGCTGCTGGATACGCTCGGCAATCGCTGCGCCTC 69
Oy 172 CCTTCAGGACCGAGGTTCACCCCTCCACCATCGCAGATCTCAACACCATGTTGTCTGCAAC 231
Db 68 CCGTTTGGCCCCCGCGACCTCCCGCCTCTTTGCTGCTATCTTTGCGTTCTTCGCGTCGAC 9
Oy 232 AC 233
Db 8 GC 7

Search completed: March 19, 2000, 08:56:37
Job time: 579 sec

Attie

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GACTTCAGTGTCTCTCCATCCAGGAGGCGGAGTGGGCTATGGGCTCTGGGCTGCCCC 65
 DB 305 GAGCGGGGCTCAGGCGCCTCCGAGCGCTGGGGCCCTGTGGACAGGCGCGGTG 246
 QY 66 TTGTCTCTCTTGACCTCTCTGGGAGTGCATGGAACAGGCGCGGG 114
 DB 245 CGGGCGGCGGCTGTGGGCTCTCTGGGCGGCGGCGGCTACAGGAGCGGG 197

RESULT 13

T18068/c

ID T18068 standard; cDNA; 597 BP.

AC T18068; (first entry)

DE Human ApoE4 cDNA.

KW Alzheimer disease; beta-amyloid peptide precursor; APP;

KW antisense; hyperlipoproteinaemia; cardiovascular disease;

KW gene therapy; apolipoprotein; ApoE4; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT cds

FT 1..597

FT /*tag= a

FT /transl_except= pos:505..507;_aa:Ala

FT /transl_except= pos:553..555;_aa:Ser

FT CA2126787-A.

PD 28-DEC-1995.

PF 27-JUN-1994; 126787.

PR 27-JUN-1994; CA-126787.

PA (PRED/) PREDDIE R E.

PI Bergmann JE, Preddie RE;

DR WPI; 96-151776/16.

DR P-PSDB; R92115.

PT Proteins and nucleic acids associated with Alzheimer's disease -

PT used to develop products for diagnosis of Alzheimer's disease and

PT related conditions, hyperlipoproteinaemia or cardiovascular disease

PS Claim 9; Fig 3A; 67pp; English.

CC 2 Open reading frames (T18068 and T18069) identified in the

CC apolipoprotein E (ApoE) antisense sequence code for proteins

CC ApoE4L (R92113) and ApoE4L1 (R92114), respectively. Both ORFs

CC are transcribed from an upstream regulatory region (T18074).

CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to

CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of

CC the significance of these and other sequences (see also T18065-77)

CC in the aetiology of Alzheimer's disease (AD) provides means for

CC diagnosing AD and related diseases, for the design of therapeutic

CC reagents (e.g. ribozymes or antibodies) and potentially for gene

CC therapy.

CC SQ Sequence 597 BP; 63 A; 241 C; 196 G; 97 T;

Query Match 12.7%; Score 29; DB 1; Length 597;
 Best Local Similarity 54.1%; Pred. No. 3.4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GACTTCAGTGTCTCTCCATCCAGGAGGCGGAGTGGGCTATGGGCTCTGGGCTGCCCC 65
 DB 305 GAGCGGGGCTCAGGCGCCTCCGAGCGCTGGGGCCCTGTGGACAGGCGCGGTG 246
 QY 66 TTGTCTCTCTTGACCTCTCTGGGAGTGCATGGAACAGGCGCGGG 114
 DB 245 CGGGCGGCGGCTGTGGGCTCTCTGGGCGGCGGCGGCTACAGGAGCGGG 197

RESULT 14

T18070/c

ID T18070 standard; cDNA; 936 BP.

AC T18070;

DT 22-MAY-1996 (first entry)

DE Human ApoE4Lx2 cDNA.
 KW Alzheimer disease; beta-amyloid peptide precursor; APP;
 KW antisense; hyperlipoproteinaemia; cardiovascular disease;
 KW gene therapy; apolipoprotein; ApoE4Lx2; ss.
 OS Homo sapiens.

FH Key Location/Qualifiers

FT cds

FT 1..597

FT /*tag= a

FT /transl_except= pos:505..507;_aa:Ala

FT /transl_except= pos:553..555;_aa:Ser

FT CA2126787-A.

PD 28-DEC-1995.

PF 27-JUN-1994; 126787.

PR 27-JUN-1994; CA-126787.

PA (PRED/) PREDDIE R E.

PI Bergmann JE, Preddie RE;

DR WPI; 96-151776/16.

DR P-PSDB; R92115.

PT Proteins and nucleic acids associated with Alzheimer's disease -

PT used to develop products for diagnosis of Alzheimer's disease and

PT related conditions, hyperlipoproteinaemia or cardiovascular disease

PS Claim 13; Fig 6A; 67pp; English.

CC 2 Open reading frames (T18068 and T18069) identified in the

CC apolipoprotein E (ApoE) antisense sequence code for proteins

CC ApoE4L (R92113) and ApoE4L1 (R92114), respectively. Both ORFs

CC are transcribed from an upstream regulatory region (T18074).

CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to

CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of

CC the significance of these and other sequences (see also T18065-77)

CC in the aetiology of Alzheimer's disease (AD) provides means for

CC diagnosing AD and related diseases, for the design of therapeutic

CC reagents (e.g. ribozymes or antibodies) and potentially for gene

CC therapy.

CC SQ Sequence 936 BP; 149 A; 317 C; 336 G; 134 T;

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 6 GACTTCAGTGTCTCTCCATCCAGGAGGCGGAGTGGGCTATGGGCTCTGGGCTGCCCC 65
 DB 305 GAGCGGGGCTCAGGCGCCTCCGAGCGCTGGGGCCCTGTGGACAGGCGCGGTG 246
 QY 66 TTGTCTCTCTTGACCTCTCTGGGAGTGCATGGAACAGGCGCGGG 114
 DB 245 CGGGCGGCGGCTGTGGGCTCTCTGGGCGGCGGCGGCTACAGGAGCGGG 197

RESULT 15

T06957

ID T06957 standard; cDNA to mRNA; 1157 BP.

AC T06957;

DT 19-JUN-1996 (first entry)

DE Human apolipoprotein-E (ApoE) cDNA.

KW Recombinant; human; apolipoprotein-E; ApoE; insect cells; larva;

KW Manduca sexta; Autographa californica nuclear polyhedrosis virus;

KW haemolymph; lipid complex; biologically active; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT misc_feature

FT 62..119

FT /*tag= a

FT note= "misc_signal"

FT 62..1015

FT /*tag= b

FT cds

FT US5472858-A.

PD 05-DEC-1995.

PF 04-JUN-1991; 709949.

PR 04-JUN-1991; US-709949.

PA (WISC) WISCONSIN ALDAMI RES FOUND.

PI Attie AD, Beckage NE, Gretch DG, Sturley SL;

DR WPI; 96-029812/03.

DR P-PSDB; R86791.





PT Prodn. of recombinant apo-lipoprotein E in insects - by infecting
PT Manduca sexta larvae with recombinant Autographica californica
PT nuclear polyhedrosis baculovirus vector.
PS Disclosure: Columns 11-14: 10pp; English.
CC Recombinant human apolipoprotein-E (ApoE) (R86791) can be produced
CC by preparing a genetic construct (contg. an ApoE-encoding sequence,
CC e.g. 106957, and flanking regulatory sequences enabling the protein
CC to be expressed in insect cells), which is then introduced into a
CC Manduca sexta larva (using a recombinant Autographica californica
CC nuclear polyhedrosis virus) and recovering the protein from the
CC haemolymph of the larval host. The ApoE produced is in a form
CC sufficiently complexed with lipids to be biologically active, which
CC cannot be achieved in insect cell cultures, and can therefore be
CC used in therapeutic applications.
CC Sequence 1157 Bp; 212 A; 370 C; 426 G; 149 T;
SQ

Query Match 12.7%; Score 29; DB 1; Length 1157;
Best Local Similarity 54.1%; Pred. No. 4.3;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 6 GACTTCAGTGTCTCTCCATCCAGAGCGCGAGTGGCCACTATGGGGTCTGGGCTGCCCC 65
DB 626 GAGCGCGGCTCAGCGCCATCCGCGAGCGCTGGGGCCCCCTGGTGAACAGGCGCGGTG 685
QY 66 TTGTCCTCTCTTGACCTCTCTGGCAGCTCACATGGAACAGGCGCGG 114
DB 686 CGGCGCGGCTCTGTGGCTCTCTGGCGCCGCGCTACAGGAGCGGG 734

Search completed: March 20, 2000, 18:37:54
Job time: 8528 sec

Bonaldo

Seq primer: -40UP from Gibco
High quality sequence stop: 395.
Location/Qualifiers
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2408041"
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia) digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 97 a 105 c 117 g 84 t 1 others
ORIGIN
Query Match 74.88; Score 175; DB 61; Length 404;
Best Local Similarity 99.48; Pred. No. 6.6e-41; Indels 0; Gaps 0;
Matches 175; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 59 CAGGGCCGGGTGATCTTGGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 118
Db 311 CAGGGCCGGGTGATCTTGGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCT 252
QY 119 ATGAGTCCAGCTCTCTGGAATGCTGAAAGCTGCTCCTCCTCCATCTCCCTTCAG 178
Db 251 ATGAGTCCAGCTCTCTGGAATGCTGAAAGCTGCTCCTCCTCCATCTCCCTTCAG 192
QY 179 GGACACGCTCAGCTCCACCATCGAAGATCTCAACACCATGTTGTCTGCAACACA 234
Db 191 GGACACGCTCAGCTCCACCATCGAAGATCTCAACACCATGTTGTCTGCAACACA 136

RESULT 2
AI136523/c
LOCUS AI136523 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-nq-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
UI-R-C2p-nq-e-02-0-UI 3', mRNA sequence.
ACCESSION AI136523
VERSION AI136523.1 GI:3637300
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 328)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150222.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wesg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to identify it as a clone from the normalized adult Lung library. cDNA library preparation: M. Fatima Bonaldo, Ph.D. Clone distribution: clones will be available through Research

AI018775 ov32e12.x
AI829206 wk76f01.x
AI010074 EST04525
T41524 10105 Lambd
AA091822 ot45c07.s
AW013553 11P028k W
A0718755 HS-5511.B
AI757641 ea33d07.Y
A0740265 HS-5505.A
T41535 10116 Lambd
T41523 10104 Lambd
A0062745 A0062745
AI981365 pat.pK005
AQ290553 nxbx0037C
AQ373909 RPI111-14
M89011 CEL14A7 Chr
R90433 16788 Lambd
T44883 8146 Lambd
T76643 11421 Lambd
H76456 18161 Lambd
T44613 7876 Lambd
AI765674 w183D02.x
AL042852 DXF2P434H
AW025169 wu94g07.x
T52782 wa79f01.r1
Z45258 HSC2KH091.n
T41525 10106 Lambd
AI803834 tp30412.x
AI442028 sb66f12.y
A0076246 A0076246
T36119 EST96907.Hu
R15181 yf86c11.r1
T59783 yf66e04.r1
R34436 y957g06.r1
AA04444 zw37c01.r
AA43415 zw87408.r
AI219495 qp24a03.x
AI692905 wd42f12.x

ALIGNMENTS

RESULT 1
AI857998/c
LOCUS AI857998 404 bp mRNA EST 26-AUG-1999
DEFINITION wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
mRNA sequence.
ACCESSION AI857998
VERSION AI857998.1 GI:5511614
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 404)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://lnc.lanl.gov/bbrp/image/image.html


```

#authors      Paquin, B.; Lang, B.F.
#journal      J. Mol. Biol. (1996) 255:688-701
#title        The mitochondrial DNA of Allomyces macrogynus: the complete
#             genomic sequence from an ancestral fungus.
#cross-references MUID:96226032
#accession    S63638
#status       nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues     1-382 #label PAQ
#cross-references EMBL:041388; NID:g1236403; PIDN:AAC9221.1;
                PDB:g1236404
#note         the nucleotide sequence was submitted to the EMBL Data
                Library, November 1995

```

GENETICS

cob
cob
mitochondrion
67/3: 137/3; 143/3; 164/1; 200/2; 252/3
#superfamily cytochrome b; cytochrome b homology; cytochrome
b8 homology; plastocytin--plastocyanin reductase 17K
protein homology
electron transfer; heme; iron; plastocyanin; oxidative
phosphorylation; oxidoreductase; respiratory chain

CLASSIFICATION

#domain cytochrome b homology label CBH\
#domain cytochrome b6 homology #label CS6\
#domain plastocytin--plastocyanin reductase 17K protein
homology #label 17K\
#binding_site heme iron, low potential (His) (axial
ligands) #status predicted\
#binding_site heme iron, high potential (His) (axial
ligands) #status predicted

KEYWORDS

#length 382 #molecular-weight 43467 #checksum 6973

FEATURE

10-340
10-210
222-340

82,197

96,202

SUMMARY

Query Match	Best Local Similarity	Score 93:	DB 2:	Length 382;
Matches	17;	Conservative	15;	Mismatches 16; Indels 3; Gaps 3;
8	PVLSANSLFADSPSPNTTYLWNGFSLG-LCLVQIVTGVTLAMHAPS	57		
22	PGMTVLQK-LKEPSL-TNSGYESSFLEKCLLHLHPSTSVTTHARS	70		

```

3
RESULT
ENTRY
  #type complete
  #translation regulator GCD6 - yeast (Saccharomyces cerevisiae)
  #guanine nucleotide exchange factor chain GCD6; protein
  #YD8142B.03; protein YDR211w; translation initiation factor
  #eIF-2B homolog
  #formal name Saccharomyces cerevisiae
  #28-May-1993 #sequence_revision 03-May-1994 #text_change
  #06-Feb-1998
  #A48156; S61578; S30776
  #A48156
  #Bushman, J.L.; Asuru, A.I.; Matts, R.L.; Hinnebusch, A.G.
  #Mol. Cell. Biol. (1993) 13:1920-1932
  #Evidence that GCD6 and GCD7, translational regulators of
  #GCN4, are subunits of the guanine nucleotide exchange
  #factor for eIF-2 in Saccharomyces cerevisiae.
  #cross-references NUID:93180841
  #accession A48156
  #molecule_type DNA
  #residues 1-712 #label BUS
  #cross-references EMBL:L07115; NID:g171572; PID:g171574
  #note
  #sequence extracted from NCBI backbone (NCBIN:126018,
  #NCBIP:126021)

```

REFERENCE
#authors Oliver, K.; Harris, D.
#submission submitted to the ENML Data Library, December 1995
#accession S61576
#molecule_type DNA
#features 1-712 #label OLI
#cross-references ENML:z68195; NID:g1122341; PID:e213795; PID:g1122344;
#cross-references MIPS:YDR211;

```
##experimental_source strain AB972
GENETICS
#gene SGD:GCD6
##cross-references SGD:S0002619; MIPS:YDR211w
#map_position 4R
translation regulation
#length 712 #molecular-weight 81160 #checksum 142
KEYWORDS
SUMMARY
Query Match 16.5%; Score 92; DB 2; Length 712;
Best Local Similarity 39.5%; Pred. No. 1.23e+00;
Matches 17; Conservative 8; Mismatches 14; Indels 4; Gaps 4;
12 LGN-HGKSDMDYEDLQAVVLTDS-YETREMPLTAVKPCLL 52
14 LGSSRGTPGNTLOLKESFTVNSSTESFELLE-KL-CLL 54
```

```

4
RESULT
ENTRY
  TITLE
  ORGANISM
  DATE
  26-Aug-1999
ACCESSIONS
REFERENCE
  #authors
  #journal
  #title
  #cross-references MVID:97128324
  #contants
  #accession
  #status
  #molecule_type mRNA
  #residues 1-252 ##label CHE
  #cross-references GB:L76632; NID:G1220536; PID:G1220537
  This protein is rich in cysteine content and play a part of the
  flowers defense mechanisms against pathogens.
CLASSIFICATION
  #superfamily thaumatin I
  #domain signal sequence #status predicted #label SIG\
  #product osmotin-like protein #status predicted #label
  MAT
  #length 252 #molecular-weight 27665 #checksum 2939
SUMMARY
  1-23
  24-252

```

Db 10 LPXSLFTLLSQSTNPNEITL 33
 III III: III: I: I I I
 Ov 5 LPXITLNYCASHGTCOMTOL 28

```

RESULT      5
ENTRY       573757      #type complete
TITLE       hypothetical protein F11_orf879 - Mycoplasma pneumoniae (ATCC
ORGANISM    29342      #normal_name Mycoplasma pneumoniae
VARIETY     ATCC 29342  #SGC3)
DATE        27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change
              17-Jul-1998
ACCESSIONS  573757
REFERENCE   573737
AUTHORS     Himmelsreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li,
              B.C.; Herrmann, R.
JOURNAL     Nucleic Acids Res. (1996) 24:4420-4449
TITLE       Complete sequence analysis of the genome of the bacterium
              Mycoplasma pneumoniae.
CROSS-REFERENCES  MWID:97105885
ACCESSION   573757
STATUS      preliminary; nucleic acid sequence not shown;
              translation not shown

```


		FEATURES	source
*		gap of unknown length	
*	47023	51524: contig of 4502 bp in length	
*		gap of unknown length	
*	51525	53246: contig of 1722 bp in length	
*		gap of unknown length	
*	53247	56192: contig of 2946 bp in length	
*		gap of unknown length	
*	56193	58896: contig of 2704 bp in length	
*		gap of unknown length	
*	58897	61546: contig of 2650 bp in length	
*		gap of unknown length	
*	61547	65673: contig of 4127 bp in length.	
		Location/Qualifiers	
		1. .65673	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="16"	
		/clone="RPCI-11.529K1"	
	16932 a	16632 c 16335 g 15686 t	87 others

Query Match	8.3%	Score 34.6	DB 42	Length 65673
Best Local Similarity	54.3%	Prd. No. 7.2		
Matches 70	Conservative 0	Mismatches 59	Indels 0	Gaps 0
190	TGAAGAAGCTGTGGCTCTCTCCATCTCCCTTCAGGGACCAAGGTCACCTCCACCATGC 249			
39815	TGAANAATCTCCATCATCTCCGCTCCCTCTCTGGAGGGGGAGCCCTCGACCCCTTC 39756			
250	AGATCTCAACACCATCTGTGTCTGCAACATCATCAGACGCATTGAAGCCTGTGCTCTCTT 309			
39755	TAACTTCAAGCCCAAAATCTCTGTGATGCACCGGCTCTCTTTTGGAAATTAATACCTTT 39696			
310	GGCCCGGGC 318			
39695	GATCTGGC 39687			

RESULT	12
SDSJ144C9/c	
LOCUS	HSDJ144C9 7732 bp DNA PRI 22-NOV-1999
SOURCE	Human DNA sequence from clone RPI-144C9 on chromosome 1p34.3-36.11,
DEFINITION	complete sequence.
ACCESSION	AL036774
VERSION	AL036774.9 GI:6465842
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutharia; Primates; Catarrhini; Hominoidea; Homo.
REPERENCE	1 (bases 1 to 7732)
AUTHORS	Hall R

TITLE	COMMENT
<p>Direct Submission Submitted (22-Nov-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk</p>	<p>On Nov 22, 1999 this sequence version replaced GI:6138775. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.</p> <p>This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.</p> <p>The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information</p>

on the WormPep database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>
RPI-144C9 is from the library RPI-1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-144C9.

[illegible]

```

disc: 780692    1275   1339
       matches: GGGG smAQIAGAGC
BASE COUNT      21006  18906  18290  19125  1
ORIGIN
Query Match           3   8.3%   Score 34.4   DB 10   Length 77322
Best Local Similarity = 63.1%   Pred. No 8.4
Matches 53; Conservative 0 Mismatches 31 Indels 0 Gaps
ay 14 TGTCCTCCATCCCAGAGCGCAGTGGCCACTATGGGTCTGGGGCTGGCCCCCTGTGCCTCT
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
bb 31300 TGTTCCTCCCTCCCAAGATCCCTTTTGTTGGTAGTATGTTTCAGGATGCCACCAACCAC 31261

```

yy	74	CTCTTGACCTCCTTGGCACTCA	97
b	31240	CTCATGATACCTTCAGGCACACA	31217
RESULT 13		D86631/c	
	LOCUS	D86631	3626 bp DNA ROD
	DEFINITION	Mus sp. DNA for oxytocin receptor, partial cds.	07-FEB-1999
	ACCESSION	D86631	
	VERSION	D86631.1	GI:1902965
	KEYWORDS	oxytocin receptor.	
	SOURCE	Mus sp. (strain:J29SV), ES cell DNA.	
	ORGANISM	Mus sp.	
		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
		Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
		1 (bases 1 to 3626)	
		Kubota, Y.	
REFERENCE		Direct Submission	
AUTHORS			
TITLE			
JOURNAL			
		Submitted (26-JUL-1996) to the DDBJ/EMBL/GenBank databases. Yasue	

Kubota, Osaka University Medical School, Department of Obstetrics and Gynecology; 2-2 Yamadaoka, Suita, Osaka 565, Japan (E-mail: kubota@gyne.med.osaka-u.ac.jp, Tel: 81-6-879-3351, Fax: 81-6-879-3359)

2 (bases 1 to 3626)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Sai, J. F. and Murata, Y.

Structure and Expression of the Mouse Oxytocin Receptor Gene Unpublished (1996)

3 (sites)

Kubota, Y., Kimura, T., Hashimoto, K., Tokugawa, Y., Nobunaga, K., Azuma, C., Sai, J. F. and Murata, Y.

Structure and expression of the mouse oxytocin receptor gene Mol. Cell. Endocrinol. 124 (1-2), 25-32 (1996)

97179034

Location/Qualifiers

1. 3626

/organism="Mus sp."

/strain="129SV"

/db_xref="taxon:10095"

/cell_type="ES cell"

1. 1530

/number=1

1799..2240

/number=1

2241..2393

/number=2

2394..2485

/number=2

2486..3547

/number=3

2629..3546

/codon_start=1

/product="oxytocin receptor"

/protein_id="BAAL18996.1"

/db_xref="GI:1902966"

/translation="MEGTAAANWSTIELDGVPPEAGNITAGPRNEALARVEA VZLILFLALSNACVLIALRTTRKHSRLFFPMKHLSDLVAVVAVFVLPQLMDIT FRFGDPLCLRLVLYQVGMFASVILLMLSLDRCLAIQPLRSRRRTDRVLAV WGLCVASVPQVPIHLSREVAGVQCDWAVEIQPMGPRVYVITLAVIVPVLVLA CYGLISFKTKQNLKRLKATAAAGAGSDAGAGRAALRAVSSVKLISKAKIRVKT FIVLAFIVCTVPTFFVQMVSDVNAPE"

3548..3626

/number=3

Intron

BASE COUNT 782 a 921 c 1131 g 792 t

ORIGIN

Query Match 8.2%; Score 34.2; DB 12; Length 3626;

Best Local Similarity 60.0%; Pred. No. 7.8; 38; Indels 0; Gaps 0;

Matches 57; Conservative

QY 152 ACAATTCTCCTATGAGTCCAGCTTCTCTGGAAATTCGTTGAAAGCTCTCTCTCTCTC 211

DB 2085 AAAAGCTTACTACGGGTACATCTCTCTGGGGTCCGCCAAAAGCCATCTCTTCTTCC 2026

QY 212 CATCTCCCTTCAGGACGACGCTCACCTCCACCA 246

DB 2025 CTCTGCTTCCAGCTACGACCCGCTCCCA 1991

RESULT 14

HSJ858B16 36676 bp DNA PRI 23-NOV-1999

LOCUS

DEFINITION

Human DNA sequence from clone 858B16 on chromosome 22. Contains KIA00542 gene and a gene for a novel protein similar to hamster PSSC (phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65). Contains ESTs, GSSs and a putative CpG island, complete sequence.

AL096768

AL096768.7 GI:5596770

LOCUS

DEFINITION

WTG: CpG island; KIA00542; Phosphatidylserine Decarboxylase Proenzyme; PSSC.

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 14:42:21 ; Search time 479.47 Seconds
(without alignments)
-1139.903 Million cell updates/sec

Title: US-09-092-296-3
Perfect score: 180
Sequence: 1 CAGGAGCCAGTGGCCACTA.....CTTGAAAGCTCGCTCCT 180

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pri.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sv.*
- 14: gb_un.*
- 15: gb_v1.*
- 16: gb_v2.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sv.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_v1.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: gb_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*
- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	90.8	50.4	192650	10	AB023048	AB023048 Homo sapi
2	90.8	50.4	200000	10	AF000511	AF000511 Homo sapi
3	90.8	50.4	47323	11	AC005937	AC005937 Homo sapi
4	33.4	18.6	97580	11	HS91N18	AL031594 Human DNA
5	32.2	17.9	36676	11	HS785B16	AL096768 Human DNA
6	31.2	17.3	177241	11	HS402G11	AL022328 Human DNA
7	31.2	17.3	303319	42	AC006927	AC006927 Homo sapi
8	31	17.2	97339	44	AB011039	AB011039 Homo sapi
9	30.8	17.1	5280	9	AB011114	AB011114 Homo sapi
10	30.8	17.1	168986	40	AC008013	AC008013 Homo sapi
11	30.8	17.1	110000	43	AC007623_0	AC007623 Homo sapi
12	30.8	17.1	179622	43	AC009533	AC009533 Homo sapi
13	30.4	16.9	1091	7	SAGAPDHR	X04301 Mustard mRN
14	30.4	16.9	1301	12	AF059567	AF059567 Mus muscu
15	30.4	16.9	1098	12	MMCARH	Y10320 M.musculus
16	30.4	16.9	1669	12	MMCARHOM	Y11929 M.musculus
17	30.4	16.9	1131	12	MMINK4B01	U66084 Mus musculu
18	30.4	16.9	1206	12	MMU90715	U90715 Mus musculu
19	30.2	16.8	2803	2	TNU93354	U93354 Thermotoga
20	30	16.7	751	7	ATRNABBC1	X75162 A.thaliana
21	30	16.7	85109	8	ATT2J13	AL132967 Arabidops
22	30	16.7	4341	9	HUMHMG2A	M83665 Human high
23	30	16.7	124990	11	HS434P1	Z97056 Human DNA S
24	30	16.7	159125	44	AC015713	AC015713 Homo sapi
25	29.8	16.6	42259	6	LBPHIGLE	X98106 Lactobacill
26	29.8	16.6	6042	6	LBPHIGP49	X90510 Lactobacill
27	29.8	16.6	2160	9	D84361	D84361 Human mRNA
28	29.8	16.6	154791	45	AC015478	AC015478 Homo sapi
29	29.6	16.4	1691	4	CHKC1PA2M	J00838 Chicken typ
30	29.6	16.4	1683	4	GGC010	V00390 Messenger f
31	29.6	16.4	151600	9	AP000346	AP000346 Homo sapi
32	29.6	16.4	1000	9	HSTALIDNA	X81027 H.sapiens t
33	29.6	16.4	168933	10	HS210I18	AL035405 Human DNA
34	29.6	16.4	188016	33	AC007030	AC007030 Homo sapi
35	29.6	16.4	204617	35	AC006074	AC006074 Drosophil
36	29.4	16.3	1755	10	HS297017	Z97017 Homo sapien
37	29.4	16.3	149698	11	HS76C18	Z98755 Human DNA s
38	29.4	16.3	135545	12	MMHC213L3	AF109905 Mus muscu
39	29.4	16.3	168609	41	AC009297	AC009297 Homo sapi
40	29.4	16.3	291288	42	AC008878	AC008878 Homo sapi
41	29.4	16.3	106687	43	AC015892	AC015892 Mus muscu
42	29.4	16.3	208079	44	AC013391	AC013391 Homo sapi
43	29.4	16.3	259307	45	AC015657	AC015657 Mus muscu
44	29.4	16.3	31311	45	AC017192	AC017192 Drosophil
45	29.2	16.2	94179	42	AC010028	AC010028 Drosophil

ALIGNMENTS

RESULT 1
AB023048
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:53L9, complete sequence.
ACCESSION AB023048
VERSION AB023048.1 GI:5672603
KEYWORDS HTG.
SOURCE Homo sapiens cell_line:978SK DNA, clone:53L9.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (sites)
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikawa,E., Iwata,K., Tomizawa,M., Okuki,N., Kuwano,Y.,
Matanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahram,S. and Inoko,H.
Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
20027539

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 192650)
Shina,T. and Takishima,N.
Direct Submission
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tshina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)

FEATURES

Location/Qualifiers

1..192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SK"
/chromosome="6"
/clone="5319"
/map="6p21.3"

BASE COUNT 49862 a 44743 c 45833 g 52212 t

ORIGIN

Query Match 50.4%; Score 90.8; DB 10; Length 192650;
Best Local Similarity 97.9%; Pred. No. 3.6e-19;
Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 78 CAGGCGGGTATGACTTGCACCTGAAGCTGAAGGCTTTCTGCAAGTCTCTCT 137
|||||
Db 177979 CAGGCGGGTATGACTTGCACCTGAAGCTGAAGGCTTTCTGCAAGTCTCTCT 178038
|||||

Oy 138 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGCT 171
|||||

Db 178039 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGT 178072
|||||

RESULT 2

AP000511 200000 bp DNA PRI 28-SEP-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
DEFINITION section 10/20.
AP000511
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

GI:5926698

Homo sapiens DNA.

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Homiidae; Homo.
1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
Published Only in Database (1999) In press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)

This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS

COMMENT

project of JST

Japan Science and Technology Corporation (JST)
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or send email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES

source

1..200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"

STS

108774..108864
/standard_name="D6S2088"
/note="SHGC-12985:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735268"

STS

115648..115906
/standard_name="D6S1898"
/note="WI-9418:The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"

STS

complement(115665..115896)
/note="Cda0vh10:The location is between each flanking site
of PCR primers."
/db_xref="GDB:443382"

STS

142030..142159
/standard_name="D6S2203"
/note="SHGC-16870:The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"

STS

complement(142085..142322)
/note="RH18132:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"

STS

complement(142479..142740)
/standard_name="D6S1851"
/note="SHGC-10808:The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"

STS

complement(164525..164665)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"

STS

complement(169209..169574)
/standard_name="D6S952"
/note="UT5233:The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"

BASE COUNT

52605 a 47531 c 49366 g 50498 t

ORIGIN

Query Match 50.4%; Score 90.8; DB 10; Length 200000;
Best Local Similarity 97.9%; Pred. No. 3.6e-19;
Matches 92; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 78 CAGGCGGGTATGACTTGCACCTGAAGCTGAAGGCTTTCTGCAAGTCTCTCT 137
|||||

Db 110306 CAGGCGGGTATGACTTGCACCTGAAGGCTTTCTGCAAGTCTCTCT 110365
|||||

Oy 138 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGCT 171
|||||

Db 110366 ATGAGTCAGCTTCTCGAATGCTTGAAGAAGT 110399
|||||

RESULT 3

AC005937
LOCUS Homo sapiens clone UMG:370M23.002 from 6p21, complete sequence.
DEFINITION AC005937
ACCESSION AC005937
VERSION AC005937.1 GI:3845393
KEYWORDS HTG.

SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 47323)

REFERENCE
AUTHORS

Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and
Geraghty, D.E.

TITLE

Large scale sequence analysis of the human MHC class I region

JOURNAL
REMARK

Unpublished (1998)
Fred Hutchinson Cancer Research Center
The Clinical Research Division
1100 Fairview Ave. N., P.O. Box 19024
Seattle, WA 98109-1024

REFERENCE
AUTHORS

2 (bases 1 to 47323)
Geraghty, D.E. and Olson, M.V.

TITLE

Direct Submission
Submitted (05-NOV-1998) Human Genome Center, University of
Washington, Box 352145, Seattle, WA 98195, USA

JOURNAL
REMARK

University of Washington Human Genome Center
Box 352145 Seattle, WA 98195
Contact: Daniel E. Geraghty (geraghty@fhcr.org)

COMMENT

Overlapping Sequences:
5': UWGC:370m23.013 (GenBank Accession: AC005530)
3': UWGC:Y67c112 (GenBank Accession: AC004211)

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Double stranded (DS) coverage: 75.5%

DS or two chemistry coverage: 98.9%

Single stranded regions: 3

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
Mapping. Comparison of the experimentally derived map digest
fragments with sequence-predicted fragments is given below.
Small fragments below a variable cutoff (approximately 400-600bp)
are not mapped and hence do not appear in the table. There are no
significant remaining discrepancies between the experimental and
predicted values. Uniquely ordered fragment groups are separated
by dashed lines.

BglII	Map	Seq	HindIII	Map	Seq	NsiI	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00		
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00		
2171.50	2147.00	7268.78	7196.00				
2560.20	2531.00	10085.80	9992.00				
4335.42	4269.00	11212.78	11131.00				
2698.62	2628.00						
1927.50	1887.00						
3130.46	3090.00						
2166.69	2129.00						
2044.67	2005.00						

FEATURES

Location/Qualifiers

1..47323

SOURCE

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21"
/sub_clone="UWGC:370M23.002"
/clone_lib="Research Genetics BAC Library"
3647..3932
/rpt_family="Alu"
complement(4999..5277)
/rpt_family="Alu"
6285..6572
/rpt_family="Alu"
complement(6972..7050)
/rpt_family="MLT1"
7286..7584
/rpt_family="Alu"
complement(8164..8609)
/rpt_family="Alu"
complement(21287..21895)
/rpt_family="Alu"
22715..22957
/rpt_family="Alu"
25510..25802
/rpt_family="Alu"
27835..28010
/rpt_family="MER20"
31295..31594
/rpt_family="Alu"
33515..33767
/rpt_family="Alu"
34223..34290
/rpt_family="MIR"
37372..37648
/rpt_family="Alu"
38526..38700
/rpt_family="MER3"
39583..40010
/rpt_family="Alu"
40046..40156
/rpt_family="Alu"
43194..43372
/rpt_family="MER5"
43325
/note="clonal variation with 3' overlapping clone"
44149
/note="clonal variation with 3' overlapping clone"
44451
/note="clonal variation with 3' overlapping clone"
44537
/note="clonal variation with 3' overlapping clone"
44814
/note="clonal variation with 3' overlapping clone"
44965
/note="clonal variation with 3' overlapping clone"
45760
/note="clonal variation with 3' overlapping clone"
45900
/note="clonal variation with 3' overlapping clone"
46851
/note="clonal variation with 3' overlapping clone"
46859
/note="clonal variation with 3' overlapping clone"
47032
/note="clonal variation with 3' overlapping clone"
47240..47256
/note="clonal variation with 3' overlapping clone"
/note="clonal variation with 3' overlapping clone -
insertion of 17bp repeat"
BASE COUNT 11556 a 11489 c 12284 g 11994 t
ORIGIN

Query Match 50.4%; Score 90.8; DB 11; Length 47323;
Best Local Similarity 97.9%; Pred. NO. 3.4e-19;

```

Matches 92: Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 78 CAGGGCGGGTATGACTTTCACACTGAAGCTGAAGAGTCTTTCTGACAAAGTTCTCTCT 137
|||||
Db 35465 CAGGGCGGGTATGACTTTCACACTGAAGCTGAAGAGTCTTTCTGACAAATTCCTCTCT 35524
|||||
QY 138 ATGAGTCCAGCTCTCGAATTCCTTGAAGAGCT 171
|||||
Db 35525 ATGAGTCCAGCTCTCTCGAATTCCTTGAAGAGGT 35558
|||||

RESULT 4
HS591N18
LOCUS
DEFINITION
    HS591N18 97580 bp DNA PRI 23-NOV-1999
    Human DNA sequence from clone 591N18 on chromosome 22q13.1-13.2.
    Contains a COX6B (Cytochrome C Oxidase subunit VIb (EC 1.9.3.1))
    pseudogene, ESTs, GSSs and two putative CpG islands, complete
    sequence.
ACCESSION
    AL031594.9 GI:5050980
KEYWORDS
    HTG; COX6B; CpG Island; Cytochrome C Oxidase.
SOURCE
    human.
ORGANISM
    Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
    Eutheria; Primates; Catarrhini; Hominidae; Homo.
    1 (bases 1 to 97580)
    Steward,C.
    Direct Submission
    Submitted (25-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
    CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
    Requests: clonerequest@sanger.ac.uk
    On Jun 11, 1999 this sequence version replaced gi:4914529.
    During sequence assembly data is compared from overlapping clones.
    Where differences are found these are annotated as variations
    corresponding to the overlapping clone name. Note that the
    variation annotation may not be found in the sequence submission
    only a small overlap as described above.
    The following abbreviations are used to associate primary accession
    numbers given in the feature table with their source databases:
    Em: EMBL; Sw: SWISSPROT; Tr: TREMBL
    This sequence has been finished according to sequence map criteria
    as follows. An attempt is made to resolve all sequencing problems,
    such as compressions and repeats, but not necessarily within known
    annotated human repeat sequence elements (e.g. Alu). Where the
    sequence is ambiguous, there is an annotation using the 'unsure'
    feature key.
    This sequence was generated from part of bacterial clone contigs of
    human chromosome 22, constructed by the Sanger Centre Chromosome 22
    Mapping Group. Further information can be found at
    http://www.sanger.ac.uk/HGP/Chr22
    591N18 is from the library RPC14 constructed at the Roswell Park
    Cancer Institute by the group of Pieter de Jong. For further
    details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2
    IMPORTANT: This sequence is not the entire insert of clone 591N18.
    It may be shorter because we only sequence overlapping sections
    once, or longer because we arrange for a small overlap between
    neighbouring submissions.
    The true left end of clone 229A8 (286090) is at 97475 in this
    sequence. The true right end of clone 1042K10 (AL022238) is at
    34197 in this sequence.
    Location/Qualifiers
        1..97580
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="22"
            /map="q13.1-13.2"
            /clone="RP4-591N18"
            /clone_lib="RPC1-4"
            1..227
                /note="AluY repeat: matches 75..300 of consensus"
                996..1104
                    /note="FLAM_C repeat: matches 1..108 of consensus"

FEATURES
    source
        1147..1220
            /note="AluY/FRAM repeat: matches 217..294 of consensus"
        1225..1501
            /note="AluY repeat: matches 12..282 of consensus"
        1552..1864
            /note="AluY repeat: matches 1..311 of consensus"
        1869..2134
            /note="AluY repeat: matches 21..275 of consensus"
        2942..3234
            /note="AluSc repeat: matches 4..295 of consensus"
        4854..5144
            /note="AluY repeat: matches 1..308 of consensus"
        6310..6584
            /note="match: GSS AQ038173"
        6506..6575
            /note="MER3 repeat: matches 134..209 of consensus"
        6585..6665
            /note="FLAM_A repeat: matches 27..116 of consensus"
        6673..6694
            /note="MER3 repeat: matches 148..167 of consensus"
        6695..6988
            /note="AluSx repeat: matches 1..290 of consensus"
        6989..7068
            /note="MER3 repeat: matches 58..148 of consensus"
        7497..7761
            /note="AluY repeat: matches 1..301 of consensus"
        7845..8136
            /note="AluSx repeat: matches 1..299 of consensus"
        8467..8776
            /note="AluY repeat: matches 1..311 of consensus"
        9209..9505
            /note="AluSx repeat: matches 1..295 of consensus"
        11780..12007
            /note="L1ME1 repeat: matches 5465..5691 of consensus"
        12018..12387
            /note="HERV23 repeat: matches 1184..1573 of consensus"
        12454..12544
            /note="L1ME1 repeat: matches 5679..5763 of consensus"
        12545..12843
            /note="AluY repeat: matches 1..297 of consensus"
        12844..13179
            /note="L1ME1 repeat: matches 5763..6159 of consensus"
        13240..13373
            /note="AluY repeat: matches 1..136 of consensus"
        13408..13612
            /note="AluSg/x repeat: matches 96..293 of consensus"
        13625..13924
            /note="AluSx repeat: matches 1..301 of consensus"
        13935..14230
            /note="AluSx repeat: matches 1..297 of consensus"
        14237..14466
            /note="MIR repeat: matches 11..261 of consensus"
        14729..14785
            /note="MIR repeat: matches 202..261 of consensus"
        15147..15467
            /note="L2 repeat: matches 2370..2690 of consensus"
        15625..15910
            /note="AluSg repeat: matches 1..296 of consensus"
        16000..16300
            /note="AluSg repeat: matches 1..291 of consensus"
        16422..16732
            /note="AluY repeat: matches 1..311 of consensus"
        16733..16813
            /note="L2 repeat: matches 2667..2746 of consensus"
            complement(16814..17293)
        /note="match: GSSs AQ377982 AQ380087"
        16907..17100
            /note="MIR repeat: matches 2..197 of consensus"
        18063..18376
            /note="AluY repeat: matches 1..307 of consensus"
        18462..18537
            /note="738 copies 2 mer tt 78% conserved"
        18540..19522

```

misc_feature /note="SVA repeat: matches 2. .954 of consensus"
19061. .20976
/note="Random repeat. Some base discrepancies edited.
Assembly is consistent with Restriction Digest."
19115. .20959
misc_feature /note="CpG island"
/evidence=not_experimental
repeat_region 19540. .20001
/note="SVA repeat: matches 521. .954 of consensus"
repeat_region 20019. .20360
/note="SVA repeat: matches 521. .845 of consensus"
repeat_region 20361. .20854
/note="SVA repeat: matches 519. .994 of consensus"
misc_feature complement(20855. .20945)
/note="match: GSS AQ473058"
repeat_region 21493. .21658
/note="AluX repeat: matches 1. .165 of consensus"
repeat_region 21793. .22086
/note="AluX repeat: matches 1. .297 of consensus"
repeat_region 22379. .22534
/note="FRAM repeat: matches 1. .162 of consensus"
repeat_region 23688. .24049
/note="MLX1A1 repeat: matches 1. .365 of consensus"
repeat_region 24100. .24408
/note="AluYb8 repeat: matches 1. .305 of consensus"
repeat_region 25227. .25536
/note="AluYb repeat: matches 1. .310 of consensus"
repeat_region 25556. .25773
/note="AluYb repeat: matches 58. .297 of consensus"
repeat_region 25834. .26147
/note="AluSg repeat: matches 1. .307 of consensus"
repeat_region 26148. .26192
/note="MIR repeat: matches 187. .230 of consensus"
repeat_region 26377. .26688
/note="AluX repeat: matches 2. .311 of consensus"
repeat_region 26884. .27010
/note="FLAMC repeat: matches 1. .133 of consensus"
repeat_region 30237. .30344
/note="MIR repeat: matches 103. .206 of consensus"
repeat_region 31467. .31777
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region 31963. .32080
/note="AluYb repeat: matches 1. .118 of consensus"
repeat_region 32118. .32408
/note="AluYb repeat: matches 1. .292 of consensus"
repeat_region 32623. .32927
/note="AluYb repeat: matches 1. .300 of consensus"
repeat_region 33031. .33354
/note="AluSg repeat: matches 1. .304 of consensus"
repeat_region 33381. .33519
/note="L2 repeat: matches 2572. .2709 of consensus"
repeat_region 33560. .33777
/note="AluX repeat: matches 1. .219 of consensus"
repeat_region 34284. .34597
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 35062. .35370
/note="AluJ repeat: matches 1. .309 of consensus"
repeat_region 35600. .35753
/note="MIR repeat: matches 2. .230 of consensus"
repeat_region 35782. .35854
/note="MER21B repeat: matches 720. .790 of consensus"
repeat_region 35912. .36626
/note="MER21B repeat: matches 1. .713 of consensus"
repeat_region 36660. .36969
/note="Alu repeat: matches 1. .308 of consensus"
repeat_region 37152. .37273
/note="AluSg repeat: matches 185. .302 of consensus"

Query Match 18.6; Score 33.4; DB 11; Length 97580;
Best Local Similarity 57.08; Pred. No. 1.6;
Matches 61; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
50.TTGCACCTCACATGACACGGCCGGGTATGACTTTGCACTGAAGAGTCTT 119

Db 2606 TTCCCTCCATCCCCAGGTTTAGTAAAGACTTTTCAAAAGCAAGGTAGGTGAAC 2665
QY 120 TTCTGACAAAGTTCTCTATGAGTCAGCTCCCTGGAAATTCCTTGA 166
Db 2666 TGTAGAAGATTCATGTTATAGTACCGCACTTGTATTCTCGAA 2712

RESULT 5
HSJ858B16 36676 bp DNA PRI 23-NOV-1999
LOCUS
DEFINITION
Human DNA sequence from clone 858B16 on chromosome 22. Contains the KIAA0542 gene and a gene for a novel protein similar to hamster PSSC (Phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65). Contains ESTs, GSSs and a putative CpG island, complete sequence.
AL096768
ACCESSION
AL096768.7 GI:5596770
VERSION
HTG; CpG Island; KIAA0542; Phosphatidylserine Decarboxylase Proenzyme; PSSC.
KEYWORDS
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Eukaryota; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 36676)
Barlow, K.
TITLE
Direct Submission
JOURNAL
Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT
On Jul 27, 1999 this sequence version replaced gi:5579004.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
858B16 is from the library RPCI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://pacpac.med.buffalo.edu/VECTOR: pcypac2
IMPORTANT: This sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone 694E4 (AL031255) is at 36572 in this sequence.
Location/Qualifiers
1. .36676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone_lib="RPCI-5"
/clone="RP5-858B16"
join(-273, 454, 1413, 1488, 1978, 2073, 2538, 2614, 3997, 4099, 5606, 5702, 8812, 8972, 9435, 9509, 10811, 10885, 11023, 11199, 11271, 11524, 12177, 12948, 14414, 14546, 14669, 14747, 15784, 15895, 15983, 16217)
/gene="dJ858B16.1"
/note="match: ESTs: Em:R53362 Em:278408 Em:AA521233...


```

repeat_region 24501..24747
/note="L1MEC repeat: matches 1501..1751 of consensus"
repeat_region 24978..25194
/note="L1MEC repeat: matches 272..492 of consensus"
repeat_region 25933..26246
/note="AluY repeat: matches 1..309 of consensus"
repeat_region 27926..28225
/note="AluX repeat: matches 2..302 of consensus"
repeat_region 28694..28997
/note="AluSg repeat: matches 1..304 of consensus"
repeat_region 29001..29119
/note="AluJ/FLAM repeat: matches 3..133 of consensus"
repeat_region 29859..30170
/note="L1MEC repeat: matches 5189..5489 of consensus"
repeat_region 30171..30449
/note="AluX repeat: matches 24..304 of consensus"
repeat_region 30450..30556
/note="L1MEC repeat: matches 5086..5189 of consensus"
repeat_region 30567..30813
/note="L2 repeat: matches 1079..1303 of consensus"
repeat_region 30975..31047
/note="L1MC/D repeat: matches 5325..5397 of consensus"
repeat_region 31048..31362
/note="AluY repeat: matches 1..303 of consensus"
repeat_region 31363..31990
/note="L1MC/D repeat: matches 4762..5325 of consensus"
repeat_region 31997..32351
/note="MLT/Al repeat: matches 16..365 of consensus"
repeat_region 32365..32907
/note="L1MC/D repeat: matches 4184..4755 of consensus"
repeat_region 32908..33208
/note="AluX repeat: matches 1..301 of consensus"
repeat_region 33209..33464
/note="L1MC/D repeat: matches 3936..4184 of consensus"
repeat_region 33468..33637
/note="FRAM repeat: matches 1..175 of consensus"
repeat_region 33650..33707
/note="L1MEC repeat: matches 2384..2075 of consensus"
repeat_region 33708..33998
/note="AluX repeat: matches 1..288 of consensus"
repeat_region 33999..34144
/note="L1MEC repeat: matches 2231..2385 of consensus"
repeat_region 34781..34916
/note="4 copies 34 mer 87% conserved"
repeat_region 34792..34910
/note="7 copies 17 mer 74% conserved"
repeat_region 35157..35466
/note="AluSg repeat: matches 1..310 of consensus"
repeat_region 35626..35922
/note="AluSg repeat: matches 1..297 of consensus"
repeat_region 35932..36042
/note="L2 repeat: matches 2643..2750 of consensus"
repeat_region 36532..37622
/note="L1MEC repeat: matches 255..1397 of consensus"
repeat_region 37623..37663
/note="AluSg repeat: matches 208..248 of consensus"
repeat_region 37706..37865
/note="AluSg/x repeat: matches 127..286 of consensus"
repeat_region 37881..38081
/note="L1MC5 repeat: matches 7541..7757 of consensus"
repeat_region 38099..38311

```

Query Match
Best Local Similarity 57.0%; Score 31.2; DB 11; Length 177241;
Matches 57; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

```

QY 16 CACTATGGGTGTCGGCTGCCCTTCCTCTTGACCTCTTGCGAGTCACATGG 75
|||||
Db 109678 CACGAGTCGCCGAGCGGCTCTCCGCGCTCTGAGCCCATGCCACCCACATGG 109619
|||||

QY 76 AACAGGCCGGGTATGACTTTGCACTGAAGCTGAAGAG 115
|||||
Db 109618 AACATGCTGGGTGGACTGCAGGATGGAGCTGGGGCAG 109579
|||||

```

AC006927 303319 bp DNA HTG 21-OCT-1999
Homo sapiens chromosome 12p13.1-17.1-21.3 clone RPC111-444J21, ***
SEQUENCING IN PROGRESS ***, 139 unordered pieces.
AC006927 GI:6091655
HTG: HTGS_PHASE1.
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 303319)
Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodoty,D.M., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondrjewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,K.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sugang,R., Tabor,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabba,M., Watlington,S.,
Wenstorf,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 303319)
Worley,K.C.
Submitted (26-FEB-1999) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030 USA
On Oct 21, 1999 this sequence version replaced gi:5882346.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 139 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 830: contig of 830 bp in length
* 831 1650: contig of 830 bp in length
* 1651 2498: contig of 838 bp in length
* 2499 3178: contig of 680 bp in length
* 3179 3829: contig of 651 bp in length
* 3830 4616: contig of 787 bp in length
* 4617 5463: contig of 847 bp in length
* 5464 6330: contig of 867 bp in length
* 6331 7341: contig of 1011 bp in length
* 7342 7832: contig of 491 bp in length
* 7833 8684: contig of 852 bp in length
* 8685 9411: contig of 727 bp in length
* 9412 10220: contig of 809 bp in length
* 10221 11025: contig of 805 bp in length
* 11026 11817: contig of 792 bp in length
* 11818 12571: contig of 754 bp in length
* 12572 13368: contig of 797 bp in length
* 13369 14328: contig of 860 bp in length
* 14329 15013: contig of 785 bp in length
* 15014 15730: contig of 717 bp in length
* 15731 16617: contig of 887 bp in length
* 16618 17335: contig of 718 bp in length
* 17336 17992: contig of 657 bp in length
* 17993 18287: contig of 295 bp in length


```

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE
1 (bases 1 to 97339)
AUTHORS
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 97339)
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Hearford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,N.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meidrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-SEP-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 7, 1999 this sequence version replaced gi:6001999.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RN/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2935
Center clone name: 7_H_18
-----
* NOTE: This record contains 109 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 873: contig of 873 bp in length
* 874 1755: contig of 882 bp in length
* 1756 2659: contig of 904 bp in length
* 2660 3548: contig of 889 bp in length
* 3549 4385: contig of 837 bp in length
* 4386 5261: contig of 876 bp in length
* 5262 6130: contig of 869 bp in length
* 6131 6989: contig of 859 bp in length
* 6990 7867: contig of 878 bp in length
* 7868 8755: contig of 888 bp in length
* 8756 9645: contig of 890 bp in length
* 9646 10550: contig of 905 bp in length
*
11434: contig of 884 bp in length
gap of unknown length
12328: contig of 894 bp in length
gap of unknown length
13231: contig of 903 bp in length
gap of unknown length
14127: contig of 896 bp in length
gap of unknown length
14998: contig of 871 bp in length
gap of unknown length
15866: contig of 868 bp in length
gap of unknown length
16752: contig of 866 bp in length
gap of unknown length
17666: contig of 914 bp in length
gap of unknown length
18549: contig of 883 bp in length
gap of unknown length
19434: contig of 885 bp in length
gap of unknown length
20319: contig of 885 bp in length
gap of unknown length
21222: contig of 903 bp in length
gap of unknown length
22101: contig of 879 bp in length
gap of unknown length
23010: contig of 909 bp in length
gap of unknown length
23881: contig of 871 bp in length
gap of unknown length
24775: contig of 894 bp in length
gap of unknown length
25667: contig of 892 bp in length
gap of unknown length
26533: contig of 866 bp in length
gap of unknown length
27420: contig of 887 bp in length
gap of unknown length
28313: contig of 893 bp in length
gap of unknown length
29205: contig of 892 bp in length
gap of unknown length
30104: contig of 899 bp in length
gap of unknown length
31008: contig of 904 bp in length
gap of unknown length
31897: contig of 889 bp in length
gap of unknown length
32805: contig of 908 bp in length
gap of unknown length
33697: contig of 892 bp in length
gap of unknown length
34581: contig of 884 bp in length
gap of unknown length
35473: contig of 892 bp in length
gap of unknown length
36343: contig of 870 bp in length
gap of unknown length
37239: contig of 896 bp in length
gap of unknown length
38129: contig of 890 bp in length
gap of unknown length
39012: contig of 883 bp in length
gap of unknown length
39921: contig of 909 bp in length
gap of unknown length
40806: contig of 885 bp in length
gap of unknown length
41670: contig of 864 bp in length
gap of unknown length
42550: contig of 880 bp in length
gap of unknown length
43411: contig of 861 bp in length

```


KEYWORDS SOURCE ORGANISM

HTG.
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 168986)

Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Korrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
Kondrjewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (10-JUL-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (28-AUG-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

4 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (02-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

5 (bases 1 to 168986)

Worley,K.C.

Direct Submission

Submitted (09-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Aug 28, 1999 this sequence version replaced gi:5739545.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality

standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

----- Summary Statistics -----
Contig length: 168986
Phrap values in estimate: 165936
Average error rate (BCM-Phrap estimate): 0.000156145
Fraction of Phrap values less than 40 : 0.03878
Number of consensus changing edits: 14
Number of N's in consensus : 0

----- Consensus changing edits -----
Position Original+Context Edited+Context
4287 cctaaacccc(n)cagctctcca cctaaacccc(a)cagctctcca
5004 ctggctctca(n)tcgccgcctc ctggctctca(g)tcgccgcctc
5018 ccgctctgca(n)gggggctctc ccgctctgca(g)gggggctctc
5050 atggggggccc(n)aagagcctgg atggggggccc(g)aagagcctgg
86008 catagagccca(n)ccttgccctg catagagccca(g)ccttgccctg
102306 gaccatcgtg(n)aagccccagg gaccatcgtg(g)aagccccagg
102546 gggtaacccc(n)cttgccctt gggtaacccc(c)cttgccctt
102547 ggtaacccc(n)cttgccctt ggtaacccc(a)cttgccctt
109591 ttgtgtttg(n)ttgttttt ttgtgtttg(t)ttgttttt
109596 ttgtgtttg(n)ttgttttt ttgtgtttg(t)ttgttttt
155478 tgcatgtg(n)gtgcatagc tgcatgtg(c)gtgcatagc
165391 aggaagga(n)aaggaagga aggaagga(g)aaggaagga
165392 aggaagga(n)aaggaagga aggaagga(g)aaggaagga
165993 aactcaaaaa(n)gggtgcagca aactcaaaaa(g)gggtgcagca

----- Distribution of Quality < 40 Bases -----
1000| * * * * *
900| * * * * *
800| * * * * *
700| * * * * *
600| * * * * *
500| * * * * *
400| * * * * *
300| * * * * *
200| * * * * *
100| * * * * *
0| * * * * *

5 10 15 20 25 30 35 40
Phrap Value Range

FEATURES

Version: 1.01 xfgc.

Location/Qualifiers

source

1..168986
/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="12p12-21.3-21.8"

/clone="RPC111-551L14"

507..655

/rpt_family="MIR"

complement(725..1166)

/rpt_family="MER110A"

981..1329

/note="Region: 0146106.s1 Homo sapiens cDNA, AA917071"

complement(1338..1761)

/rpt_family="MLTIC"

2100..2251

/rpt_family="MER112"

3343..3431

repeat_region

* 194817 194873: gap of unknown length
* 194874 196660: contig of 1787 bp in length
* 196661 196717: gap of unknown length
* 196718 198486: contig of 1769 bp in length
* 198487 198543: gap of unknown length
* 198544 200297: contig of 1754 bp in length
* 200298 200354: gap of unknown length
* 200355 202066: contig of 1712 bp in length
* 202067 202123: gap of unknown length
* 202124 203824: contig of 1701 bp in length
* 203825 203881: gap of unknown length
* 203882 205545: contig of 1664 bp in length
* 205546 205602: gap of unknown length
* 205603 207253: contig of 1651 bp in length
* 207254 207310: gap of unknown length
* 207311 208900: contig of 1590 bp in length
* 208901 208957: gap of unknown length
* 208958 210510: contig of 1533 bp in length
* 210511 210567: gap of unknown length
* 210568 212117: contig of 1550 bp in length
* 212118 212174: gap of unknown length
* 212175 213721: contig of 1547 bp in length
* 213722 213778: gap of unknown length
* 213779 215323: contig of 1545 bp in length
* 215324 215380: gap of unknown length
* 215381 216948: gap of unknown length
* 216949 218452: contig of 1504 bp in length
* 218453 218509: gap of unknown length
* 218510 219991: contig of 1482 bp in length
* 219992 220048: gap of unknown length
* 220049 221520: contig of 1472 bp in length
* 221521 221577: gap of unknown length
* 221578 223033: contig of 1456 bp in length
* 223034 224519: contig of 1429 bp in length
* 224520 224576: gap of unknown length
* 224577 225977: contig of 1401 bp in length
* 225978 226034: gap of unknown length
* 226035 227416: contig of 1382 bp in length
* 227417 227473: gap of unknown length
* 227474 228834: contig of 1361 bp in length
* 228835 228891: gap of unknown length
* 228892 230288: contig of 1337 bp in length
* 230289 230285: gap of unknown length
* 230290 231603: contig of 1318 bp in length
* 231604 231660: gap of unknown length
* 231661 232977: contig of 1317 bp in length
* 232978 233034: gap of unknown length
* 233035 234335: contig of 1301 bp in length
* 234336 234392: gap of unknown length
* 234393 235688: contig of 1296 bp in length
* 235689 235745: gap of unknown length
* 235746 236860: contig of 1115 bp in length
* 236861 236917: gap of unknown length
* 236918 238201: contig of 1284 bp in length
* 238202 238258: gap of unknown length
* 238259 239538: contig of 1280 bp in length
* 239539 239595: gap of unknown length
* 239596 240860: contig of 1265 bp in length
* 240861 240917: gap of unknown length
* 240918 242182: contig of 1265 bp in length
* 242183 242239: gap of unknown length
* 242240 243497: contig of 1258 bp in length
* 243498 243554: gap of unknown length
* 243555 244805: contig of 1251 bp in length
* 244806 244862: gap of unknown length
* 244863 246101: contig of 1239 bp in length
* 246102 246158: gap of unknown length
* 246159 247381: contig of 1223 bp in length
* 247382 247438: gap of unknown length
* 247439 248638: contig of 1200 bp in length
* 248639 248695: gap of unknown length

* 248696 249893: contig of 1198 bp in length
* 249894 249950: gap of unknown length
* 249951 251148: contig of 1198 bp in length
* 251149 251205: gap of unknown length
* 251206 252402: contig of 1197 bp in length
* 252403 252459: gap of unknown length
* 252460 253655: contig of 1196 bp in length
* 253656 253712: gap of unknown length
* 253713 254901: contig of 1189 bp in length
* 254902 254958: gap of unknown length
* 254959 256146: contig of 1188 bp in length
* 256147 256203: gap of unknown length
* 256204 256954: contig of 751 bp in length
* 256955 257011: gap of unknown length
* 257012 258197: contig of 1186 bp in length
* 258198 258254: gap of unknown length
* 258255 258436: contig of 1182 bp in length
* 258437 259493: gap of unknown length
* 259494 260675: contig of 1182 bp in length
* 260676 260732: gap of unknown length
* 260733 261911: contig of 1179 bp in length
* 261912 261968: gap of unknown length
* 261969 263137: contig of 1169 bp in length
* 263138 263194: gap of unknown length
* 263195 264350: contig of 1156 bp in length
* 264351 264407: gap of unknown length
* 264408 265560: contig of 1153 bp in length
* 265561 265617: gap of unknown length
* 265618 266751: contig of 1134 bp in length
* 266752 266808: gap of unknown length
* 266809 267938: contig of 1130 bp in length
* 267939 267995: gap of unknown length
* 267996 269117: contig of 1122 bp in length
* 269118 269174: gap of unknown length
* 269175 270289: contig of 1115 bp in length
* 270290 270346: gap of unknown length
* 270347 271458: contig of 1112 bp in length
* 271459 271515: gap of unknown length
* 271516 272627: contig of 1112 bp in length
* 272628 272684: gap of unknown length
* 272685 273794: contig of 1110 bp in length
* 273795 273851: gap of unknown length
* 273852 274952: contig of 1101 bp in length
* 274953 275009: gap of unknown length
* 275010 276109: contig of 1100 bp in length
* 276110 276166: gap of unknown length

Query Match 17.1% Score 30.8; DB 43; Length 110000;
Best Local Similarity 63.5%; Pred. No. 12;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 2 AGGAGCCAGTGGCCACTATGGGTCTGGGCTGGCCCTGTGCTCTCTCTGACCTCTT 61
DB 75932 AGGCGCTGAGGACCCACTGTGCTGGGCTGGCCATCTGGCCATCTGTGACCATCTCC 75991
QY 62 GGCAGCTCACATGG 75
DB 75992 AGGGGCTCTCAGGG 76005

RESULT 12
AC009533/c
LOCUS
DEFINITION Homo sapiens clone RP11-22B23, *** SEQUENCING IN PROGRESS ***, 26
AC009533 DNA HTG 19-NOV-1999
ACCESSION AC009533
VERSION AC009533.6 GI:6143848
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 179622)

Db 772 TTCTGCTGTCACAGGAACG 791

Search completed: March 21, 2000, 14:44:47
Job time: 3902 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 20, 2000, 18:37:54 ; Search time 70.18 Seconds

(without alignments)
641.701 Million cell updates/sec

Title: US-09-092-296-3
Perfect score: 180
Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAAGCTGCTCCT 180

Scoring table: IDENTITY_NUC

Searched: 311585 segs, 125096042 residues

Database : N_Geneseg_36.*

Word size : 0

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	178.4	99.1	439	1	V84366	Human stomach carc
2	32	17.8	128	1	V28846	Mouse cxsackievir
3	30.4	16.9	580	1	T02964	Cell-cycle regulat
4	30.4	16.9	1301	1	V50430	Mouse cxsackievir
5	29.8	16.6	291	1	V86365	EST clone AR34. Ne
6	28.8	16.0	1785	1	T66386	Human brain-specif
7	28.6	15.9	1554	1	Q71243	Serine protease fo
8	27.6	15.3	3088	1	T05628	ADP ribosylation f
9	27.4	15.2	2259	1	Q05106	Sequence encoding
10	27.2	15.1	2122	1	T83188	Wnt-10b gene. Expr
11	26.8	14.9	2703	1	Q83695	New TGF-beta fami
12	26.8	14.9	2341	1	T31601	Cartilage-derived
13	26.8	14.9	2703	1	T59729	Human bone morphog
14	26.8	14.9	2703	1	T59405	DNA encoding huma
15	26.8	14.9	2703	1	T61412	DNA encoding huma
16	26.8	14.9	2703	1	T69695	DNA encoding huma
17	26.8	14.9	2703	1	T98191	DNA for human MP5
18	26.8	14.9	2703	1	T88340	Human MP52 cDNA. M
19	26.6	14.8	9487	1	Q79440	Hepatitis C virus
20	26.6	14.8	2782	1	V71227	C-Delta-1 gene (al
21	26.6	14.8	834	1	X39963	Prostate cancer as
22	26.4	14.7	1275	1	T78903	FE65 cDNA. Peptide
23	26.4	14.7	1275	1	V36874	Nucleotide sequenc
24	26.4	14.7	1329	1	V33199	Secreted protein F
25	26.4	14.7	378	1	V87853	EST clone FB78. Ne
26	26.4	14.7	2539	1	V08822	Gene No. 12 encodi
27	26.4	14.7	50000	1	X23517	Human kidney amino
28	26.2	14.6	348	1	V86243	EST clone AA299. N
29	26	14.4	2883	1	T58898	C-Delta-1 gene (al
30	26	14.4	2088	1	T58897	C-Delta-1 gene. Ne
31	25.8	14.3	2319	1	Q73728	Chicken liver alph
32	25.8	14.3	2334	1	T66121	Chicken alpha-N-ac
33	25.8	14.3	2319	1	T60489	Chicken liver alph
34	25.6	14.2	944	1	V91561	BBC6 gene for regu
35	25.6	14.2	817	1	V37118	DNA sequence used
36	25.6	14.2	809	1	V37133	DNA sequence used
37	25.6	14.2	405	1	V88069	EST clone FW13. Ne
38	25.6	14.2	510	1	X19506	Degenerate lipocal
39	25.4	14.1	252	1	Q24478	Parathyroid hormon

40 25.4 14.1 263 1 Q24479 Parathyroid hormon
41 25.4 14.1 135 1 Q33578 Downstream sequenc
42 25.4 14.1 234 1 Q35228 Gene for hPTH mute
43 25.4 14.1 237 1 Q35229 Gene for hPTH mute
44 25.4 14.1 240 1 Q35230 Gene for hPTH mute
45 25.4 14.1 243 1 Q36829 Gene for hPTH mute

ALIGNMENTS

RESULT 1
V84366
ID V84366 standard; cDNA to mRNA; 439 BP.
AC V84366;1999 (first entry)
DE Human stomach carcinoma cDNA clone HP10408.
KW Transmembrane protein; HP10408; human; stomach cancer; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 75..311
FT /tag= a
FT /note= "cDNA comprising the coding region (minus
the stop codon) is claimed (Claim 3)"
FN WO9855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PI (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI: 99-045730/04.
DR P-FSD8; W88498.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 4; Page 135; 178pp; English.
CC This is the nucleotide sequence of cDNA clone HP10408, which
CC includes a coding region (also claimed) for a novel human
CC transmembrane protein (see W88498). The clone was isolated from a
CC stomach cancer cDNA library using a signal sequence detection
CC method, and by protein synthesis by in vitro translation. The
CC encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see W84359-76) coding for 18 transmembrane proteins
CC (see W88491-508), vectors containing such polynucleotides, and
CC eukaryotic cells containing the vectors. The proteins can be
CC used as antigens or as compositions in the preparation of
CC antibodies against the proteins. The polynucleotides can be used
CC as probes for gene diagnosis, and as gene sources for gene therapy
CC and large-scale production of proteins encoded by the cDNA. The
CC host cells are used for the detection of ligands corresponding to
CC the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 439 BP; 89 A; 137 C; 109 G; 104 T;

Query Match 99.1%; Score 178.4; DB 1; Length 439;
Best Local Similarity 99.4%; Pred. No. 6.8e-50;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCTCTCTCTCTGACCTCTCT 60
DB 56 CAGGAGCGCAGTGGCCACTATGGGGTCTGGGCTGCCCTCTCTCTCTGACCTCTCT 115
QY 61 TGGCAGCTCACATGAACAGCGGGGTATGATTTGCACTGAAGCTGAAGAGTCTTT 120
DB 116 TGGCAGCTCACATGAACAGCGGGGTATGATTTGCACTGAAGCTGAAGAGTCTTT 175
QY 121 TCTGCAAGTTCCTCTATGAGTCCAGCTTCTTGGATTCGTTGAAAAGCTCTGCTCT 180
DB 176 TCTGCAAAATTCCTCTATGAGTCCAGCTTCTTGGATTCGTTGAAAAGCTCTGCTCT 235

```

RESULT 2
ID V28846 standard; DNA; 1128 BP.
AC V28846;
DE Mouse coxsackievirus and adenovirus receptor encoding DNA.
KW Mouse coxsackievirus; adenovirus; receptor; CAR; cardiac infection;
KW myocarditis; pericarditis; dilated cardiomyopathy; meningoencephalitis;
KW pancreatic infection; acute pancreatitis; gastrointestinal tract;
KW diabetes mellitus; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 1..1128
FT FT /*tag= a
FT FT /product= "coxsackievirus and adenovirus receptor"
FT FT /transl_except= (pos:1096..1098,aa:taa)
FT FT /note= "taa= a stop codon, the sequence is shown as
FT FT continuing but gets unclear"
FN W09811221-A2.
PD 19-MAR-1998; U16189.
PF 12-SEP-1997; US-026100.
PR 13-SEP-1996; US-026100.
PA (DAND ) DANA FARRER CANCER INST INC.
PI Bergelson JM, Finberg RW, Horowitz MS;
DR WPI: 98-207384/18.
DR P-PSDB; W57213.
PT DNA encoding coxsackie virus and adenovirus receptor - useful for
PT modulation of e.g. cardiac, pancreatic or gastrointestinal infection
PT Disclosure; Fig 14; 104pp; English.
PS The present sequence encodes mouse coxsackievirus and adenovirus
PS receptor (CAR). The present invention also describes: (1) a method for
PS modulating CAR expression comprising contacting the cell with an agent
PS which modulates CAR protein activity or CAR nucleic acid expression,
PS such that a cell associated activity is altered relative to a cell
PS associated activity of the cell in the absence of the agent; and (2) a
PS method for detecting the presence of CAR in a biological sample
PS comprising contacting a biological sample with an agent capable of
PS detecting CAR protein or mRNA such that the presence of CAR is detected.
PS Modulation of CAR is useful for treatment of cardiac infection, e.g.
PS myocarditis, pericarditis or dilated cardiomyopathy, or infection of the
PS central nervous system, e.g. a non-specific febrile illness or
PS meningoencephalitis, or pancreatic infection, e.g. acute pancreatitis,
PS or infection of the respiratory or gastrointestinal tract or childhood
PS onset diabetes mellitus. Probes derived from CAR nucleic acids are
PS useful for hybridisation assays, and antibodies raised against CAR
PS protein are useful for blocking CAR expression. Cell-free assays which
PS include combining CAR protein and a candidate/test compound are useful
PS in screening for drugs which interact with CAR protein.
SQ Sequence 1128 BP; 297 A; 286 C; 288 G; 257 T;

Query Match 17.8%; Score 32; DB 1; Length 1128;
Best Local Similarity 62.5%; Pred. No. 0.21;
Matches 50; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 2 AGGAGCGGAGTGGCGCTATGGGCTGCTGCGCTGCCCTTGTCTCTTGTACCTCTCTT 61
DB 712 ACATCGCGGGCGCGCTATAGGAGCGTGTGCGCTTGTCTCATCGGGGCCATCTC 771
QY 62 GGCAGCTCATGTGACAGG 81
DB 772 TTCTGCTGTACAGGAAGG 791

RESULT 3
T02964/C
ID T02964 standard; cDNA; 580 BP.
AC T02964;
DE 01-MAR-1996 (first entry)
DE Cell-cycle regulatory protein p15 cDNA.
KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
KW ss; ds.

```

```

OS Mus sp.
FH Key Location/Qualifiers
FT CDS 91..483
FT FT /*tag= a
FT PN W09528483-A1.
PD 26-OCT-1995.
PF 14-APR-1995; U04636.
PR 14-APR-1994; US-227371.
PR 25-MAY-1994; US-248812.
PR 14-SEP-1994; US-306511.
PR 29-NOV-1994; US-346147.
PA (COLD-) COLD SPRING HARBOR LAB.
PI Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
DR WPI: 95-373798/48.
DR P-PSDB; R85118.
PT New cell cycle regulating proteins bind to cyclin dependent kinase -
PT and related nucleic acids, antibodies etc., used in diagnosis and
PT therapy of abnormal cell proliferation, degeneration etc.
PS Claim 43; Page 81-82; 109pp; English.
CC CDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
CC protein p15 (R85118) was isolated from an embryonal carcinoma
CC library using a probe based on a mouse CCR p13.5 cDNA (T02965).
CC The isolated cDNA can be used: to detect mutations in CCR
CC genes that lead to cell proliferation; to breed transgenic
CC animals to study cellular disorders involving CCR allele
CC mutation/misexpression; and to correct CCR-deficient cells
CC (gene therapy).
SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;

Query Match 15.9%; Score 30.4; DB 1; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.56;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 8 GCATGTGCCATATGGGTCTGGGCTGCCCTTGTCTCTTGTACCTCTTGGCAGC 67
DB 128 CGGTGTGCCAGCGCGCTACTCTGCGCGCCCAACATGCGCTTCCCGCTGTGGC 69
QY 68 TCACATGGACAGGCGGGGTATGACTTTGCCAACTG 103
DB 68 AGRATGTGCTCTCGCGCGGTGAGATGCTTACAG 33

RESULT 4
V50430
ID V50430 standard; cDNA; 1301 BP.
AC V50430;
DT 07-DEC-1998 (first entry)
DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 157..1215
FT FT /*tag= a
FT PN W09833819-A1.
PD 06-AUG-1998.
PF 30-JAN-1998; U01724.
PR 30-JAN-1997; US-036986.
PA (UTNY ) UNIV NEW YORK STATE.
PI Philipson L, Tomko RP;
DR WPI: 98-437397/37.
DR P-PSDB; W69698.
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells
PT susceptible to transformation by adenoviral vectors in gene therapy
PS Claim 2; Page 68-70; 88pp; English.
CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
CC serves as a cellular receptor for adenoviruses of the serotypes 2
CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
CC was obtained by screening a lambda phage expression cDNA library
CC with antiserum containing antibodies specific for mouse CVB
CC binding protein p46, and was used to identify human HCAR cDNA

```

CC (see V50429). The invention also provides host cells transformed
CC with DNA molecules encoding HCAR or MCAR and methods of producing
CC the recombinant proteins or their derivatives. These proteins,
CC their extracellular domains, as well as oligopeptides (see
CC W69699-708) which bind virus, are also provided. Isolated HCAR or
CC MCAR proteins or their fragments or variants are used to prevent
CC or treat virus infections. Expressing the DNA in cells which lack
CC these viral receptors renders the cells susceptible to
CC transformation by adenoviral vectors carrying genes for gene
CC therapy.
SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;

Query Match 16.9%; Score 30.4; DB 1; Length 1301;
Best Local Similarity 61.2%; Pred. No. 0.76;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 AGGAGCGCAGTGGCCACTATGGGTCTGGGCTGCCCTTGTCCCTCTTGTACCCCTCCTT 61
DB 868 ACGATCGGGCGCGCTCATAGGACGCTGTGGCCCTTGTCTCATCGGGGCCATCCTC 927

QY 62 GCGAGCTCATGTGAACAGG 81

DB 928 TTCTGCTCATAGGAAACG 947

RESULT 5

ID V86365/c V86365 standard; cDNA; 291 BP.
AC V86365:
DT 27-APR-1999 (first entry)
DE EST clone AR34.
KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN W09845435-A2.
PS 15-OCT-1998; U06954.
PF 10-APR-1998; US-835913.
PR 10-APR-1997; US-835913.
PA (GEM) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
PI WPI: 99-070076/06.
DR New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1; Page 210; 63pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 16.6%; Score 29.8; DB 1; Length 291;
Best Local Similarity 70.2%; Pred. No. 0.69;
Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 88 TATGACTTTGCACTGAAGCTGAAGTGTCTTTTTCACAAAGTCTCCTATGATGTC 144
DB 257 TTGATTTGTACAGAGCTGAAGTATTTTCTTACACATTTCTCTTAAATTC 201

RESULT 6

ID T66986/c T66986 standard; cDNA to mRNA; 1785 BP.
AC T66986:
DT 30-JUL-1997 (first entry)
DE Human brain-specific adaptor FC99 cDNA.
KW Brain-specific adaptor; FC99; neuron; signal transduction;
KW neurological disease; Alzheimer's disease; Parkinson's disease;
KW diagnosis; therapy; tyrosine kinase; ds.
OS Homo sapiens.
PN EP-773291-A2.
PD 14-MAY-1997; 116333
PF 11-OCT-1996; 116333
PR 13-OCT-1995; JP-265988.
PR 12-DEC-1995; JP-323069.
PR 29-FEB-1996; JP-069265.
PR 24-JUL-1996; JP-212973.
PA (SUNE) SUMITOMO ELECTRIC IND CO.
PI Nakamura T;
PI WPI: 97-261318/24.
DR P-PSDB; W15256.
DR Human and rat-derived brain-specific adaptor polypeptide FC99 - is
PT involved in neuron signal transduction and is useful in diagnosis
PT and treatment of neurological diseases e.g. Alzheimer's dementia
PS Claim 4; Page 22-23; 42pp; English.
CC A cDNA clone (T66986) codes for FC99 (W15256), a novel brain-
CC specific adaptor molecule involved in neuron signalling pathways.
CC It was obt'd. from a normalised cDNA library prep'd. from human
CC cerebrium-derived mRNA. 5'-terminal sequences of plasmid inserts
CC in E. coli transformants were sequenced and translated amino acid
CC sequences were compared with a protein database for homology. Over
CC 500 plasmids were analysed, and one plasmid, designated FC99 (FC -
CC forebrain cortex), was selected. Sequencing revealed alternative,
CC in-frame initiation points of translation, providing coding
CC sequences of 1425 bases (T66983) and 1785 bases (T66986). FC99
CC nucleic acids (including antisense) and polypeptides can be used to
CC study phosphorylated tyrosine-mediated intracellular signalling
CC pathways in diagnosis and treatment of neurological diseases and
CC to test diagnosis/treatment methods or reagents
SQ Sequence 1785 BP; 412 A; 549 C; 522 G; 302 T;

Query Match 16.0%; Score 28.8; DB 1; Length 1785;
Best Local Similarity 62.5%; Pred. No. 2.9;
Matches 45; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 15 CCACATATGGGGTCTGGGCTGGCCCTTCTCTCTTGTACCCCTCTGCACGCTCACATG 74
DB 419 CCCGAGGGGGGCTGGGAGCGGCTCTGCCGGCGGCGCCCTCTCTGGCGGGCTCAIG 360

QY 75 GAACAGGGCGCG 86
DB 359 GCGGGGCGCGG 348

RESULT 7

ID Q71243 Q71243 standard; cDNA; 1554 BP.
AC Q71243:
DT 22-MAR-1995 (first entry)
DE Serine protease for fusion protein cleavage.
KW Serine protease; Factor-Xa; recognition site;
KW fusion protein cleavage; protein folding; ds.
OS Bos taurus.
PN Key
PF Key
PR cds
PR mat_peptide
PR W09418227-A.
PD 18-AUG-1994.

Location/Qualifiers
76..1554
/*tag= a
76..1551
/*tag= b

Db 64 GCCACAGGCGCCTTGGGCGACCCCTCTGGCCCATCTTGGG 15

RESULT 9

Q05106

ID Q05106 standard; DNA; 2259 BP.

AC Q05106;

DT 02-NOV-1990 (first entry)

DE Sequence encoding tumour-associated antigen GA733-1.

KE Pancreatic carcinoma; GA733-1; cancer; ds.

OS Homo sapiens.

ES Key

FH Location/Qualifiers

FT Cds 307..1275

FT /*tag= a

FT PN EP-375746-A.

FT 4-JUL-1990.

PD PD

PF 29-DEC-1989; 313687.

PR 29-DEC-1988; US-291583.

PA (WIST-) Wistar Inst.

PI Linnebach A;

DR WPI: 90-203091/27.

DR P-PSDB; R05711.

PT Tumor-associated antigen, GA733-2 -

PT Expressed in pancreatic carcinoma cells, used for tumor

PT immuno:therapy.

PS Claim 2; Fig 2A; 12pp; English.

CC The peptide is produced by tumor cells, and Abs raised to the

CC peptide may be used in diagnosis and therapy of human tumours.

SQ Sequence 2259 BP; 460 A; 648 C; 639 G; 512 T;

Query Match 15.2%; Score 27.4; DB 1; Length 2259;

Best Local Similarity 54.5%; Pred. No. 9.1;

Matches 55; Conservative 0; Mismatches 46; Indels 0; Gaps

QY 14 GCCACTATGGGGCTCTGGGGTCGCCCTTGCTCTCTTGACCTCTTGGCAGCTCACAT 73

Db 1144 GTCATCGTGGTGTGCTGGTGGCCCTCTGTCGCCGCGATGGCGTCTCTGGTGATCACCAC 1203

QY 74 GGCACAGGCGCGGTATGACTTTGCACTGAAGCTGAAGGA 114

Db 1204 CGGAGAAGTCGGGGAAGTACAGAAGGTGGAGATCAAGGA 1244

RESULT 10

T49318

ID T49318 standard; cDNA; 2122 BP.

AC T49318;

DT 30-MAR-1997 (first entry)

DE Wnt-10b gene.

KW Human; Wnt-10b; fetal bone marrow stroma cell; growth factor;

KW antibody; antisense; hematopoietic stem cell culture;

KW gene therapy; ex vivo; diagnostic; ss.

OS Homo sapiens.

FH Key

FT Location/Qualifiers

FT 5'utr 1..193

FT /*tag= a

FT cds 194..1363

FT /*tag= b

FT /product= Wnt-10b protein

FT signal_peptide 194..256

FT /*tag= c

FT mat_peptide 257..1360

FT /*tag= d

FT /product= Mature Wnt-10b (claim 19)

FT 3'utr 1364..2122

FT /*tag= e

FT W09640910-A1.

PD PD

PF 29-DEC-1996.

PF 06-JUN-1996; E02455.

PR 07-JUN-1995; US-485449.

PA (SANO) SANDOZ LTD.

PA (SANO) SANDOZ PATENT GMBH.

PA (SANO) SANDOZ-BEFLINDUNGEN VERW GMBH.
 PA (SYST-) SYSTEMIX INC.
 PI Van Den Berg DJ;
 DR WPI: 97-052317/05.
 DR P-PSDB; W08928.
 PT Expression vectors and methods for cloning N-terminal signal
 PT sequences - used for selection, expression and isolation of new
 PT signal sequences
 PS Claim 3: Page 28-30; 46pp; English.
 CC The sequence encodes human growth factor Wnt-10b, and has been
 CC isolated from a fetal bone marrow stroma cell cDNA library by PCR
 CC using consensus primers corresponding to conserved sequences within
 CC the Wnt gene family, followed by use of the amplified fragment as a
 CC probe and chromosome walking for isolation of the full-length
 CC sequence. A truncated sequence, Wnt-10b-delta, is given in 749319
 CC (claim 3). A plasmid containing the Wnt-10b cDNA has been deposited
 CC as ATCC 97208. Antibodies and antisense sequences corresponding to
 CC Wnt-10b may be used as pharmaceuticals, and Wnt-10b may be used in
 CC a culture medium for ex vivo haematopoietic stem cell culture for
 CC use e.g. in gene therapy. Fragments of the gene may be used as
 CC diagnostic probes or primers.
 SQ Sequence 2122 BP; 415 A; 604 C; 635 G; 468 T;
 Query Match 15.1%; Score 27.2; DB 1; Length 2122;
 Best Local Similarity 51.7%; Pred. No. 10;
 Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
 QY 15 CCACATGGGTCTGGGCTGCCCTTGCTTCCTCTTGTGACCTCTTGGCAGCTCACATG 74
 Db 715 CCACTCTTCTGGCCAGCCCTGGCCCTGGCTCAAGCCGCCCTGGCCAGGACACATG 774
 QY 75 GAACAGGCGCGGTATGACATTTTCAACTGAAGCTGAAGGAGTCTTTCTGACAAAGTTCTCT 134
 Db 775 GGAATGGGTGGCTGTAAACATGACATGACTTTGGAGAGAGTCTCTCGGATTTCTT 834
 RESULT 11
 Q83695
 ID Q83695 standard; DNA; 2703 BP.
 AC Q83695;
 DT 10-OCT-1995 (first entry)
 DE New TGF-beta family member - MP-52 DNA sequence.
 KW Transforming growth factor-beta family; mitogenic; differentiation;
 KW treatment; prevention; disease; bone; cartilage; connective tissue;
 KW skin; mucosa; epithelium; dental tissue; wound healing; osteoporosis;
 KW tissue regeneration; arthritis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 640..2145
 FT /*tag= a
 FT /product= MP-52 TGF-beta propeptide
 FT mat_peptide 1783..2142
 FT /*tag= b
 FT /codon_start= 1783
 PN W09504819-A.
 PD 16-FEB-1995.
 PF 09-AUG-1994; E02630.
 PR 10-AUG-1993; DE-326829.
 PR 25-MAY-1994; DE-418222.
 PR 09-JUN-1994; DE-420157.
 PA (BIOP-) BIOPHARM GRS BIOTECHNOLOGISCHEN ENTWICKL.
 PI Hotten G, Neidhardt H, Paullista M, Hoetten G;
 DR WPI: 95-090897/2.
 DR P-PSDB; R69600.
 CC New DNA encoding a new member of the TGF beta family - and
 CC related vectors, host cells etc., has mitogenic and
 CC differentiation inducing activity, e.g. for treating or
 CC preventing diseases of bone and cartilage etc.
 PS Claim 1; Page 34-35; 51pp; German.
 CC The nucleotide sequence of a novel member of the transforming growth
 CC factor-beta (TGF-beta) family named MP-52. The gene encodes a protein
 CC of 501 amino acids (AA). The protein, or at least the mature protein,

CC has mitogenic and/or differentiation inducing properties useful in
 CC the treatment or prevention of diseases of bone, cartilage, connective
 CC tissue, skin, mucosa, epithelium or dental tissue. The protein can also
 CC be used for wound healing and tissue regeneration e.g. in osteoporosis
 CC and arthritis.
 SQ Sequence 2703 BP; 602 A; 758 C; 784 G; 559 T;
 Query Match 14.9%; Score 26.8; DB 1; Length 2703;
 Best Local Similarity 64.5%; Pred. No. 15;
 Matches 40; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
 QY 13 GCCCATGATGGGTGGGCTGCCCTTGCTTCCTTGTGACCTCTTGGCAGCTCAC 72
 Db 2449 GCCCACCATTCTCTCACCTGGGCTCTTCACCTCTGGACTCTCTTAAGCACCCTCA 2508
 QY 73 TG 74
 Db 2509 GG 2510
 RESULT 12
 T31601
 ID T31601 standard; DNA; 2341 BP.
 AC T31601;
 DT 25-OCT-1996 (first entry)
 DE Cartilage-derived morphogenetic protein-1 gene.
 KW Human; cartilage-derived morphogenetic protein-1; CDMP-1;
 KW articular cartilage; chondrogenic; vulnary; implantation;
 KW chondromalacia; osteoarthritis; therapy; joint repair; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_feature 1..381
 FT /*tag= a
 FT /note= "Genomic DNA sequence"
 FT cds 265..1770
 FT /*tag= b
 FT /product= Cartilage-derived morphogenetic protein-1
 FT signal_peptide 268..321
 FT /*tag= c
 FT mat_peptide 322..1767
 FT /*tag= d
 FT /product= pro-CDMP-1 protein
 FT misc_feature 382..2341
 FT /*tag= e
 FT /note= "cDNA sequence"
 FT mat_peptide 1408..1767
 FT /*tag= f
 FT /product= Mature C-terminal domain
 FT misc_feature 1501..1620
 FT /*tag= g
 FT /note= "PCR-amplified probe used to screen cDNA
 FT library"
 PN W09614335-A1.
 PD 17-MAY-1996.
 PF 07-NOV-1994; WO-U12814.
 PR 07-NOV-1994; WO-U12814.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chang SC, Luyten FP, Moos M,
 DR WPI: 96-251714/25.
 DR P-PSDB; R95635.
 CC New purified cartilage extracts and proteins - used to stimulate the
 CC development and repair of cartilage in vivo.
 PS Claim 10; Fig 1; 34pp; English.
 CC The sequence encodes human articular cartilage-derived morphogenetic
 CC protein-1 (CDMP-1). Primers T13603-06, constructed from conserved
 CC motifs from bone morphogenetic protein, have been used in reverse
 CC transcription-polymerase chain reaction amplification of a cattle
 CC articular chondrocyte cDNA library, generating products of 120 and
 CC 280 bp, which are then used to screen a human articular cartilage
 CC cDNA library. A positive clone lacking the N-terminal region has
 CC been isolated, and used to obtain a full-length genomic clone.
 CC CDMP-1 is present in a purified cartilage extract (claimed) which

QY 58 TCACATGGACAGGCGCGGTATGACTTTGCACTG 103
Db 68 AGAAATGGTCTTCGCCGCGGTGAGATTGCTACAG 33

RESULT 2

PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/346,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91...480
PCT-US95-04636-7

Query Match 16.9%; Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.12;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 8 GCATGGCCACTATGGGCTGTGGCTGCTGCTCTTGTCTCTTTGACCTCTCTTGGCAGC 67
Db 128 GCGTGCGCAGCGCGGCTCACTGCTGCGCGCCCAACATGCCCTTGTCCCGGTCTGTGC 69
QY 68 TCACATGGACAGGCGCGGTATGACTTTGCACTG 103
Db 68 AGAAATGGTCTTCGCCGCGGTGAGATTGCTACAG 33

RESULT 3

US-08-469-486-1
; Sequence 1, Application US/08469486
; Patent No. 5739281
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Hollet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: Improved method for the refolding of
; TITLE OF INVENTION: proteins
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,486
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1554 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1551
US-08-469-486-1

Query Match 15.9%; Score 28.6; DB 2; Length 1554;
Best Local Similarity 49.0%; Pred. No. 0.77;
Matches 76; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 15 CCACATATGGGCTGTGGCTGCGCCCTTGTCTCTCTTGTGACCTCTTGGCAGCTCACATG 74
Db 71 CCACATATGGGCGGCTGTGCTCATCTCTTGTCTAGCAGCGCCCTGGCGGCTCTTGC 130
QY 75 GAACAGGCGCGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAGTCTCT 134
Db 131 GCGCGGCGGAGCGGTGTCTCTGCGCGGACCGAGCCCGTCTCTCGAGAGAGCCC 190
QY 135 CCAATGAGTCCAGCTTCTCGGAATTTGCTTGAAG 169
Db 191 GCAGGGCCCACTCATCTTGGAGGAGGTGAGCAG 225

RESULT 4

US-08-469-658-1
; Sequence 1, Application US/08469658
; Patent No. 5917018
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Hollet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.


```

RESULT 7
US-08-485-449-1
: Sequence 1, Application US/08485449
: Patent No. 5824789
: GENERAL INFORMATION:
: APPLICANT: VANDENBERG, DAVID
: TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
: TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
: TITLE OF INVENTION: THOROF
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 755 Page Mill Road
: CITY: Palo Alto
: STATE: California
: COUNTRY: USA
: ZIP: 94304-1018
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/485,449
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: KONSKI, ANTOINETTE F.
: REGISTRATION NUMBER: 34,202
: REFERENCE/DOCKET NUMBER: 20296-20035.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 813-5600
: TELEFAX: (415) 494-0792
: TELEX: 705141
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2122 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 194...1360
US-08-485-449-1

Query Match 15.1%; Score 27.2; DB 2; Length 2122;
Best Local Similarity 51.7%; Pred. No. 2.7;
Matches 62; Conservative 0; Mismatches 58; Indels 0; Gaps

QY 15 CCACATATGGGTGTGGCTGGCTTGTCTCTCTCTTGTGACCCCTTTGGCAGCTCACATG 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 715 CCATCTCTCTCCAGCCCTGGCCCTGGCTCAAGCCCGCCAGCCCGCCAGGACATG 774
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 75 GAACAGGGCGGGTATGACTTTGCACTGAAGCTGAAGAGTCTTTTCTGCAAGTTCTT 134
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 775 GGAATGGGTGGTGTAACTGATGACATGACATGTCCTTTGGAGAGAAGTTCTCTCGGGATTTCTT 834
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
US-08-288-508C-1
: Sequence 1, Application US/08288508C
: Patent No. 5994094
: GENERAL INFORMATION:
: APPLICANT: H tten, Gertrud
: APPLICANT: Neidhardt, Helge
: APPLICANT: Paulista, Michael
: TITLE OF INVENTION: NEW GROWTH/DIFFERENTIATING FACTOR OF
: TITLE OF INVENTION: THE TGF- FAMILY
: NUMBER OF SEQUENCES: 40
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nikolaide, Marmelstein, Murray & Oram LLP
: STREET: 655 Fifteenth Street N.W. Suite 330

```

```

; LENGTH: 2782 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-09-172-528-1

Query Match          14.8%; Score 26.6; DB 4; Length 2782;
Best Local Similarity 53.3%; Pred.No. 5;
Matches 56; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 75 GAACAGGCGCCGGTATGACTTTCACAACTGAAGCTGAAGGAGTGCTTTTCGACAAAGTTCTCT 134
Db 2101 GAAAGCCTCCGAGAGACGAGCAGAGACAGCTGTTCTCGCTTCTTCCCTAAGGTGGAG 2042

QY 135 CCTATGAGTCCAGCTTCTCTCGAATTGCTTTGAAAGCTCTGCGCTCC 179
Db 2041 ATTCCTGAGTCTGCTGCTTTTGAACCTGATGGACAGGCAGTGCCGCC 1997

```

RESULT 11
 US-08-222-616-32
 ; Sequence 32, Application US/08222616
 ; Patent No. 5635177
 ; GENERAL INFORMATION:
 ; APPLICANT: Bennett, Brian D.
 ; APPLICANT: Goeddel, David
 ; APPLICANT: Lee, James M.
 ; APPLICANT: Matthews, William
 ; APPLICANT: Tsai, Siao Ping
 ; APPLICANT: Wood, William I.
 ; TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST
 ; TITLE OF INVENTION: ANTIBODIES

CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,616
FILING DATE: 4-APR-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/00586
FILING DATE: 22-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/826935
FILING DATE: 22-JAN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M.
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER: 821P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 4425 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-222-616-32

Db	3163	GACAGCCTTTTCGTCTGCACACTTCTTAGACACTGAAACCGGACAGGGCCCTGTAGATGTTT	3222
Qy	101	CTGAAGCTCTGAAGAGCTCTTTCTGCACAGTTCTCTCTATGACTCAGTTCCTGCGAATG	160
Db	3223	CTGGGACTGATCAGGCGTTCCTCGTCAAGGGCCGACGGGGATTTACCTACCGGGGACTT	3282
Qy	161	CTTGAAAGCTCTGCCTCTCT	180
Db	3283	TCGTAGAAGCTGTTCCACAT	3302

```

RESULT 13
US-08-406-070-1/c
; Sequence 1, Application US/08406070
; Patent No. 5610063
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel S.
; APPLICANT: Walker, John C.
; TITLE OF INVENTION: A CDNA FOR
; TITLE OF INVENTION: alpha-N-acetyl-galactosaminidase From Gallus domesticus
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reising, Ethlington, Barnard, Perry & Milton
; STREET: P.O. Box 4390

```

```

Query Match      14.3%; Score 25.8; DB 1; Length 2334;
Best Local Similarity 67.9%; Pred. No. 8.8;
Matches 36; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY    28 TGGGTCGCCCTTGTCTCTCTTTGACCCCTCTTGCGAGCTCATCATGGAACAG 80
       || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
Db     1081 TGATCTTCCCACGTGTACAGGCTTTGCACCTCATATGCAGCTCCCATGGGGAAG 1029

RESULT 14
US-08-665-617-1/c
; Sequence 1, Application US/08665617
; Patent No. 5663316
; GENERAL INFORMATION:
; APPLICANT: Xudong, Yin
; TITLE OF INVENTION: Gene and Protein for Regulation of Cell Death

```


NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwachik & Saliwachik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,617
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saliwachik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: CL-8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
us-08-665-617-1

Query Match 14.2%; Score 25.6; DB 1; Length 944;
Best Local Similarity 54.2%; Pred. No. 6.9;
Matches 52; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
Qy 4 GAGCGCATGCCCATCTGGGCTGTGGCTGCCCTTGTCTCTCTTGTGACCTCTCTCTGG 63
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 GAGCGGAGCGCGCCGAGGGGCTGGGCTCTCCGCCATCCCTCGTCTGCTCGCTCGTCC 296
Qy 64 CAGCTCACATGACGACAGGCGCGGTATGACTTTGCA 99
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 CGCGGGGTAGGACCTGTGGCGACTCCGGATCTCCA 260

RESULT 15
us-08-689-190-1
Sequence 1, Application US/08689190
Patent No. 5714349
GENERAL INFORMATION:
APPLICANT: FUKUDA, Tsunehiko
APPLICANT: OSHIKA, Turi
APPLICANT: YAMADA, Takao
TITLE OF INVENTION: SYNTHETIC GENE CODING FOR HUMAN
TITLE OF INVENTION: PARATHYROID HORMONE
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/689,190

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/006,197
FILING DATE:
APPLICATION NUMBER: US/08/016,171
FILING DATE:
APPLICATION NUMBER: US/07/765,371
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, Gregory D
REGISTRATION NUMBER: 30,901
REFERENCE/DOCKET NUMBER: 41,288
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthesizing DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..252
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: MUTATION
LOCATION: 7, 8, 9, 12, 15, 19, 21, 33, 36, 43, 51, 58, 60,
IDENTIFICATION METHOD: S
us-08-689-190-1

Query Match 14.1%; Score 25.4; DB 2; Length 252;
Best Local Similarity 56.6%; Pred. No. 4.5;
Matches 47; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 15 CCACATATGGGGTCTGGGCTGCCCTTGTCTCTCTTGTGACCTCTCTTGGCAGCTCACATG 74
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 96 CAATTTTGTGCTTAGGTGGCCCATTTGGCTCTCTGTATGCTGTCCCAAGACCACG 155
Qy 75 GAACAGGCGCGGTATGACTTTG 97
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 156 TAAAAAGGAAGACAATGTCTTAG 178

Search completed: March 22, 2000, 01:22:37
Job time: 5641 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2000, 05:09:44 ; Search time 802.03 Seconds
(without alignments)
847.374 Million cell updates/sec

Title: US-09-092-296-3
Perfect score: 180
Sequence: 1 CAGGAGCGCAGTGGCCACTA.....CTTGAAGAAGCTCTGCTCTCT 180

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*

- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
c 1	100.4	55.8	404	61	AI857998
c 2	74.2	41.2	552	82	AQ718761
c 3	55.6	30.9	328	42	AI136523
c 4	33.4	18.6	568	100	AQ290553
c 5	32.4	18.0	537	64	AW078074
c 6	32.4	18.0	523	69	AW128453
c 7	31.4	17.4	339	21	F08745

8	31	17.2	545	23	T41524	T41524 10105 Lambd
9	30.8	17.1	427	30	AA219322	AA219322 zq17d07.s
10	30.8	17.1	589	51	AI730494	AI730494 BNLGH1691
11	30.4	16.9	332	26	W70374	W70374 mel1503.r1
12	30.4	16.9	742	27	Z78408	Z78408 HSZ78408.Hu
13	30.4	16.9	290	33	AA387480	AA387480 vc83d07.r
14	30.4	16.9	524	43	AI179983	AI179983 EST223714
15	30.4	16.9	445	87	AQ768084	AQ768084 HS_3097.B
16	30	16.7	585	63	AI1998636	AI1998636 701546395
17	29.8	16.6	505	20	M89011	M89011 CEL1447 Chr
18	29.8	16.6	401	89	AQ904556	AQ904556 GSTRc0417
19	29.6	16.4	547	23	T41523	T41523 10104 Lambd
20	29.6	16.4	573	23	T41535	T41535 10116 Lambd
21	29.4	16.3	293	37	AA668276	AA668276 ab78e11.s
22	29.4	16.3	344	37	AA716892	AA716892 vu62g09.r
23	29.4	16.3	325	37	AA717760	AA717760 vt98d09.r
24	29.4	16.3	286	38	AA762265	AA762265 vw60f06.r
25	29.4	16.3	421	39	AA839043	AA839043 vw93d06.r
26	29.4	16.3	400	46	AI463192	AI463192 vw60f06.x
27	29.4	16.3	517	61	AI839539	AI839539 UI-M-AN0-
28	29.4	16.3	530	62	AI881867	AI881867 606074D05
29	29.4	16.3	408	64	AL117984	AL117984 p9223a55
30	29.4	16.3	603	64	AW066022	AW066022 687004H08
31	29.4	16.3	453	64	AW066263	AW066263 687008G03
32	29.4	16.3	446	104	AQ327267	AQ327267 CITBI-El-
33	29.4	16.3	382	105	AQ613185	AQ613185 HS_5133-A
34	29.2	16.2	601	51	AI727080	AI727080 BNLGH1726
35	29.2	16.2	592	51	AI730002	AI730002 BNLGH1593
36	29.2	16.2	677	100	AQ316270	AQ316270 RPC111-10
37	29.2	16.2	452	105	AQ647097	AQ647097 RPC193-Dp
38	29.2	16.2	499	105	AQ647978	AQ647978 RPC193-EC
39	29	16.1	420	38	AA774353	AA774353 L30-81T3
40	29	16.1	685	103	AQ473733	AQ473733 CITBI-El-
41	28.8	16.0	266	21	R09726	R09726 Yf27809.r1
42	28.8	16.0	574	51	AI730173	AI730173 BNLGH1586
43	28.8	16.0	847	99	AQ258186	AQ258186 nbxb0019H
44	28.8	16.0	457	103	AQ500124	AQ500124 V20D2 mfn
45	28.6	15.9	536	23	T41525	T41525 10106 Lambd

ALIGNMENTS

RESULT 1	AI857998	404 bp	mRNA	EST	26-AUG-1999
LOCUS	wj69b01.x1	NCI_CGAP_Lu19	Homo sapiens	cdna clone	IMAGE:2408041 3',
DEFINITION	mRNA sequence.				
ACCESSION	AI857998				
VERSION	AI857998.1				
KEYWORDS	GI:5511614				
EST.					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
COMMENT	Tumor Gene Index Unpublished (1997) On Jun 5, 1998 this sequence version replaced gi:3189004. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cdna Library Preparation: M. Bento Soares, Ph.D. DNA Sequencing by: Greg Lennon, Ph.D. Clone distribution: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/obrp/image/image.html				

Seq primer: -40UP from Gibco	
High quality sequence stop: 395.	
Location/Qualifiers	
1. 404	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone_lib="NCI_CGAP_Lu19"	
/tissue_type="squamous cell carcinoma, poorly differentiated (4 pooled tumors, including primary and metastatic)"	
/dev_stage="adult"	
/lab_host="DH10B (phage-resistant)"	
/note="organ: lung; Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from pooled lung tumor tissue, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."	
BASE COUNT	97 a 105 c 117 g 84 t 1 others
ORIGIN	
Query Match	55.8%; Score 100.4; DB 61; Length 404;
Best Local Similarity	98.1%; Pred. No. 6.7e-21;
Matches 101; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY 78	CAGGCGCGGTGACTTTCGAACTGAAGCTGAAGGAGTCTTTCTGACAAAGTCTCTCT 137
Db	
QY 311	CAGGCGCGGTGACTTTCGAACTGAAGCTGAAGGAGTCTTTCTGACAAATCTCTCT 252
Db	
QY 138	ATGAGTCAGCTTCCTGGAATTCGTTGAAAGCTCTGCTCTCT 180
Db	
QY 251	ATGAGTCAGCTTCCTGGAATTCGTTGAAAGCTCTGCTCTCT 209
Db	
RESULT 2	
LOCUS	AQ718761 552 bp DNA GSS 13-JUL-1999
DEFINITION	HS_5511_B2_F09_17A RPC1-11 Human Male BAC Library Homo sapiens genomic clone Plate-1087 Col-18 Row=L, genomic survey sequence.
ACCESSION	AQ718761
VERSION	AQ718761.1
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 552)
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
JOURNAL	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
MEDLINE	Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
COMMENT	99380589 Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 1087 row: L column: 18

Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.
Location/Qualifiers
1...552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1087 Col-18 Row-L"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"

BASE COUNT 141 a 153 c 124 g 117 t 17 others
ORIGIN

Query Match 41.2%; Score 74.2; DB 82; Length 552;
Best Local Similarity 85.4%; Pred. No. 8.7e-13;
Matches 82; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 76 AACAGGCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTCTGACAAAGTTCTCTC 135
Db 474 AGCATGCGAGGTATGACTTTGCACTGAAGCTGAAGGAGTCTTTCTGACAAAGTTCTCTC 415

QY 136 CTATGAGTCAGCTTCCTGGAAATGCTTGAAGAAGT 171
Db 414 TTATGAGTCAGCTTCCTGGAAATGCTTGAAGAAGT 379

RESULT 3
LOCUS A1136523/c 328 bp mRNA EST 11-FEB-1999
DEFINITION UI-R-C2p-ng-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
ACCESSION A1136523
VERSION A1136523
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 328)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Jan 19, 1998 this sequence version replaced gi:2150222.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dt track served to identify it as a clone from the normalized
adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
Ph.D. Clone distribution: clones will be available through Research
Genetics
Seq primer: M13 Forward.
Location/Qualifiers
1...328
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C2p-ng-e-02-0-UI"
/clone_lib="UI-R-C2p"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library, which is a subtracted library derived from the

UI-R-CO library. The UI-R-CO library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dt track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)"

BASE COUNT 62 a 77 c 98 g 91 t
ORIGIN

Query Match 30.9%; Score 55.6; DB 42; Length 328;
Best Local Similarity 66.9%; Pred. No. 3.8e-07;
Matches 79; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 63 GCAGCTCACATGGAACAGGCGGGTATGACTTTGCAACTGAAGCTGAAGGAGTCTTTTC 122
Db 327 GCAGCTCATCTCAAGGGCCCAACAAAGTGACTTTGCAAGTAACATGACCGAGGATCTC 268

QY 123 TGACAGTTCCTCTATGATGCCAGTTCCTCGGAATGCTTGAAAGCTCTGCGCTCT 180
Db 267 AGGCCAAGACCTCCCAAGACTCGGGCTTTCTGGACATGCTCCAAAGATCTGCTCTCT 210

RESULT 4
LOCUS AQ290553 568 bp DNA GSS 03-DEC-1998
DEFINITION ndx50037C20f CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION AQ290553
VERSION AQ290553.1 GI:3951915
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 568)
Wing,R.A. and Dean,R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATAGGACTCCTATAGGG
Class: BAC ends
High quality sequence start: 35
High quality sequence stop: 317.
Location/Qualifiers
1...568
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="abxb0037C20f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"

FEATURES
source

info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.rsgen.com) (email contact: info@resgen.com) and
RessourcenZentrumPrimarDatenbank, Berlin, Germany (web address:
www.rzpd.de)

Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

email: genexpress@genetnon.ir
Single read.
Genexpress_library_idt: C; Genexpress_sequence_idt: y3c-ldb01
Seq primer: (-21)M13-universal
High quality sequence stop: 277.

single read.
Genexpress_library_id: C: Genexpress_sequence_id: y3c-ldb01
Seq primer: (-21)M13_universal
High quality sequence stop: 277.
Location/Qualifiers
1..339

```

1. .339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-ldb01"
/clone_lib="normalized infant brain cdna"
/sex="Female"
/tissue_type="total brain"
/dev_stages="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site.1: HindIII;
Site.2: NotI; sex:Female; dev_stages=3 months old;
isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Soares, Psychiatry
Dept. Columbia University, USA. Normalization method:
Bento Soares, P.N.A.S. in press"

```

Query Match 17.4%; Score 31.4; DB 21; Length 339;
Best Local Similarity 63.5%; Pred. No. 11;
Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0.

[illegible]

QY 72 ATGGAACAGGGCCG 85
||| | |||
Db 141 CTGGACCCCTGCTG 154

RESULT	8	EST
T41524	T41524	07-AUG-1995
LOCUS	545 bp mRNA	
DEFINITION	10105 Embryonic 230kD acidic ribosomal protein	

ACCESSION	T41524	sequence.
VERSION	T41524.1	GI:931185

Arabisopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE
AUTHORS
1 (bases 1 to 545)
Newman, F., deBruin, F. J., Green, P., Keegstra, K., Kende, H.
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas
Retzel, E. and Somerville, C.

Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones

QY	2	AGGAGCGCATTGGCTATGGGTCTGGCCCTTGCCTTCCTCTTGACCCCTCT 51
Db	123	ACCATCGGGCGGCCGTCATAGGACGCTGCTGGCCCTTTGTGCTCACTGGGCGCACCTC 182
QY	62	GCGAGCTCACATGGACGAG 81
Db	183	TTCTGCTGTACAGGAAGC 202
RESULT	14	
Ail179983		
LOCUS		524 bp mRNA EST 20-JAN-1999
DEFINITION		RST23714 Normalized rat spleen, Bento Soares Rattus sp. cDNA clone
VERSION		RSPO49.3; end, mRNA sequence.
KEYWORDS		Ail179983
SOURCE		Ail179983.1 GI:3730621
ORGANISM		Rattus sp. Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus;
REFERENCE		1 (bases 1 to 524) Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J., Kerlavage,A.R. and Adams,M.D.
AUTHORS		Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
TITLE		Gene Index
JOURNAL		Unpublished (1998)
COMMENT		Contact: Lee, NH The Institute for Genomic Research 9712, Medical Center Drive, Rockville, MD 20850, USA Tel: (301)-838-3529 Fax: (301)-838-0208 Email: nhlee@tigr.org Seq primer: M13-21.
FEATURES		Location/Qualifiers 1..524 /organism=Rattus sp." /db_xref= "Atcc (inhost);2035352" /db_xref= "taxon:10118" /cloned_RSCPO49" /clone_lib= "Normalized rat spleen, Bento Soares" /note= "Organ: spleen; Vector: pT73pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT		129 a 124 c 141 g 130 t
ORIGIN		
Query Match		16.9%; Score 30.4; DB 43; Length 524;
Best Local Similarity		61.2%; Pred.No. 26;
Matches		Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY	57	TCCTTGGCAGCTCATGGAACAGCGGGGTATGACTTTTCAACTGAAGCTGAAGGAT 116
Db	443	TCCTTGCCACCAGATCCAACCTTCTCGGGTTTCGCTGTGAACAAACCTTGGAGCGT 502
QY	117	CTTTTCTGACAAGTTCTCTC 136
Db	503	CTTACCTGACTAGTTCCACC 522
RESULT	15	
A0768084		
LOCUS		A0768084
DEFINITION		HS_3097_b1_Ell_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3097 Col-21 Row-J, genomic survey
VERSION		A0768084.1 GI:5646200
KEYWORDS		GSS.
SOURCE		human.
ORGANISM		Homo sapiens DNA GSS 28-JUL-1999

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 445)
Mahairas.G.G., Wallace.J.C., Smith.K., Swartzell.S., Holzman.T.,
Keller.A., Shaker.R., Furlong.J., Young.J., Zhao.S., Adams.M.D. and
Hood.L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 3097 row: J column: 21
Seq primer: M13 Reverse
Class: BAC ends
High quality sequence stop: 445.
FEATURES
 source
 1..445
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3097 Col=21 Row=J"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
 E-Coli DH10B."
BASE COUNT 139 a 114 c 70 g 121 t 1 others
ORIGIN

Query Match 16.9%; Score 30.4; DB 87; Length 445;
Best Local Similarity 67.2%; Pred. No. 24;
Matches 43; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 35 CCCCCTTGTCTCTGTGACCTCTTGGCAGCTCAGTGAACAGGCGGGGTATGACT 94
 ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 101 CCCTTTGTCTGTGATGCCCTTATTACTATTCTCATGAGACGCTCCCGCTATGACT 160
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 95 TTGC 98
 |||
Db 161 ATGC 164

Search completed: March 22, 2000, 05:09:46
Job time: 1367 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 14:39:02 ; Search time 479.47 Seconds
(without alignments)
-1450.209 Million cell updates/sec

Title: US-09-092-296-2
Perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_pl1.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_vi.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*
- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	109.4	47.8	192650	10	AB023048	AB023048 Homo sapi
2	109.4	47.8	200000	10	AF000511	AF000511 Homo sapi
3	109.4	47.8	47323	11	AC005937	AC005937 Homo sapi
c 4	34.4	15.0	77322	10	HSDJ144C9	AL096774 Human DNA
c 5	34.4	15.0	211190	44	AC016716	AC016716 Homo sapi
6	34.4	15.0	191010	45	AC016764	AC016764 Homo sapi
7	34	14.8	1801	7	OSU25430	U25430 Oryza sativ
8	34	14.8	36676	11	HSJ858B16	AL096768 Human DNA
9	34	14.8	291288	42	AC008878	AC008878 Homo sapi
10	33.8	14.8	62181	44	AC016298	AC016298 Homo sapi
c 11	33.6	14.7	128379	11	HSJ773AL18	AL049557 Human DNA
12	33.6	14.7	200000	41	AC006445	AC006445 Homo sapi
c 13	33.6	14.7	26065	43	AC014817	AC014817 Drosophil
c 14	33.4	14.6	106791	45	AC017080	AC017080 Homo sapi
15	32.6	14.2	5280	9	AB011114	AB011114 Homo sapi
c 16	32.4	14.1	97580	11	HS591N18	AL031594 Human DNA
c 17	31.8	13.9	62091	32	DMR28018	AL121814 Drosophil
c 18	31.6	13.8	2803	2	TNU93354	U93354 Thermotoga
c 19	31.6	13.8	566	7	AAALTA6	X78222 A.alternaria
c 20	31.6	13.8	570	8	AAU87806	U87806 Alternaria
c 21	31.6	13.8	158986	40	AC008013	AC008013 Homo sapi
c 22	31.6	13.8	179622	43	AC009533	AC009533 Homo sapi
23	31.4	13.7	2745	3	EEU33171	U33171 Erinaceus e
c 24	31.4	13.7	39261	10	AC000387	AC000387 Genomic s
c 25	31.4	13.7	118067	10	HS163M9	AL021970 Homo sapi
c 26	31.4	13.7	303806	33	AL133353	AL133353 Homo sapi
c 27	31.4	13.7	154791	45	AC015478	AC015478 Homo sapi
c 28	31.2	13.6	506	3	CHIRDNA	X71732 C.hircus in
c 29	31.2	13.6	177241	11	HS402G11	AL022328 Human DNA
c 30	31.2	13.6	78054	35	AC004296	AC004296 Drosophil
c 31	31.2	13.6	179532	42	AC010856	AC010856 Homo sapi
c 32	31.2	13.6	245767	42	AC011607	AC011607 Homo sapi
c 33	31.2	13.6	207043	44	ATF11111	AC011612 Homo sapi
c 34	31	13.5	103150	8	ATF11111	AL079347 Arabidops
c 35	31	13.5	125502	8	ATT4L20	AL023094 Arabidops
36	31	13.5	42244	11	AC004660	AC004660 Homo sapi
37	31	13.5	39303	11	AC005595	AC005595 Homo sapi
38	31	13.5	190837	11	HS90L6	Z97353 Human DNA s
c 39	31	13.5	727	34	MITN17	L31623 Mitrocoma c
40	31	13.5	186851	40	AC007297	AC007297 Homo sapi
41	31	13.5	97339	44	AC011039	AC011039 Homo sapi
42	31	13.5	208079	44	AC013391	AC013391 Homo sapi
43	30.8	13.4	14524	10	D86997	D86997 Homo sapien
44	30.8	13.4	46604	10	D88269	D88269 Homo sapien
c 45	30.8	13.4	100635	11	AC003104	AC003104 Homo sapi

ALIGNMENTS

RESULT 1
AB023048 192650 bp DNA PRI 20-NOV-1999
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:53L9, complete sequence.
ACCESSION AB023048
VERSION AB023048.1 GI:5672603
KEYWORDS HTG.
SOURCE Homo sapiens cell_line:978SK DNA, clone:53L9.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

project of JST
Japan Science and Technology Corporation (JST)
5-3, Yonbanchyo, Chiyoda-ku, Tokyo, 102-0081 Japan
For further information about this sequences, please visit our
sequence archive Web site (<http://www-alis.tokyo.jst.go.jp/HGS/top.html>) or
send email to webmaster@www-alis.tokyo.jst.go.jp.

FEATURES
source
1. .200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
108774. .108864
/standard_name="D6S2088"
/note="SHGC-12985:The location is between each flanking
site of PCR primers."
/db_xref="GDB:735288"
115648. .115906
/standard_name="D6S1898"
/note="WI-9418:The location is between each flanking site
of PCR primers."
/db_xref="GDB:678272"
complement(115665. .115896)
/note="CdaOv10:The location is between each flanking site
of PCR primers."
/db_xref="GDB:741215"
142030. .142139
/standard_name="D6S2203"
/note="SHGC-16870:The location is between each flanking
site of PCR primers."
/db_xref="GDB:741215"
complement(142085. .142322)
/note="RH18132:The location is between each flanking site
of PCR primers."
/db_xref="GDB:4573021"
complement(142479. .142740)
/standard_name="D6S1851"
/note="SHGC-10808:The location is between each flanking
site of PCR primers."
/db_xref="GDB:675281"
complement(164525. .164665)
/note="SHGC-3064:The location is between each flanking
site of PCR primers."
/db_xref="GDB:1234116"
complement(169209. .169574)
/standard_name="D6S952"
/note="UT5233:The location is between each flanking site
of PCR primers."
/db_xref="GDB:313481"
BASE COUNT 52605 a 47531 c 49366 g 50498 t
ORIGIN

Query Match 47.88; Score 109.4; DB 10; Length 200000;
Best Local Similarity 95.08; Pred. No. 2.9e-24;
Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCT 60
|||||
Db 109962 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCT 110021
|||||

QY 61 GCGCCCTGTCTCTCCCTTGGACCTCTTGGCAGCTCAGTGGAGCCGCGGTATGA 119
|||||
Db 110022 GCGCCCTGTCTCTCCCTTGGACCTCTTGGCAGCTCAGTGGAGCCGCGGTATGA 110080
|||||

RESULT 3
AC005937 47323 bp DNA PRI 05-NOV-1998
LOCUS
DEFINITION Homo sapiens clone UW6C:370M23.002 from 6p21, complete sequence.
AC005937
ACCESSION
AC005937.1 GI:3845393
KEYWORDS
HTG.

REFERENCE
AUTHORS
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Bahrani,S. and Inoko,H.
TITLE
Molecular dynamics of MHC genes unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
MEDLINE
20027539
2 (bases 1 to 192650)
Shina,T. and Takishima,N.
Direct Submission
TITLE
Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi
Shina, Tokai University School of Medicine, Department of
Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,
Japan (E-mail:tashina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,
Fax:81-463-94-8884)
FEATURES
source
1. .192650
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="978SK"
/chromosome="6"
/clone="53L9"
/map="6p21.3"
BASE COUNT 49862 a 44743 c 45833 g 52212 t
ORIGIN

Query Match 47.88; Score 109.4; DB 10; Length 192650;
Best Local Similarity 95.08; Pred. No. 2.9e-24;
Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCT 60
|||||
Db 177635 ACCGGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCT 177694
|||||

QY 61 GCGCCCTGTCTCTCCCTTGGACCTCTTGGCAGCTCAGTGGAGCCGCGGTATGA 119
|||||
Db 177695 GCGCCCTGTCTCTCCCTTGGACCTCTTGGCAGCTCAGTGGAGCCGCGGTATGA 177753
|||||

RESULT 2
AP000511 200000 bp DNA PRI 28-SEP-1999
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region,
section 10/20.
ACCESSION AP000511
VERSION AP000511.1 GI:5926698
KEYWORDS
SOURCE
Homo sapiens DNA.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Shina,S., Tamiya,G., Oka,A. and Inoko,H.
TITLE
Homo sapiens 2,229,818bp genomic DNA of 6p21.3 HLA class I region
JOURNAL
Published only in database (1999) in press
2 (bases 1 to 200000)
Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.
Direct Submission
TITLE
Submitted (21-SEP-1999) to the DDBJ/EMBL/GenBank databases. Mika
Hirakawa, Japan Science and Technology Corporation (JST), Advanced
Databases department; 5-3, Yonbanchyo, Chiyoda-ku, Tokyo 102-0081,
Japan (E-mail:mika@tokyo.jst.go.jp,
URL:<http://www-alis.tokyo.jst.go.jp/>, Tel:81-3-5214-8491,
Fax:81-3-5214-8470)
This sequence is conducted by Tokai University as a JST sequencing
Team.
Principal Investigator: Hidetoshi Inoko Ph.D
Phone:+81-463-93-1121, Fax:+81-463-94-8884,
The sequence is submitted by Human Genome Sequencing in ALIS

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 47323)
 AUTHORS Janer, M., Guillaudeau, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.
 TITLE Large scale sequence analysis of the human MHC class I region
 JOURNAL Unpublished (1998)
 REMARK Fred Hutchinson Cancer Research Center
 The Clinical Research Division
 1100 Fairview Ave. N., P.O. Box 19024
 Seattle, WA 98109-1024
 REFERENCE 2 (bases 1 to 47323)
 AUTHORS Geraghty, D.E. and Olson, M.V.
 TITLE Direct Submission
 JOURNAL Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
 REMARK University of Washington Human Genome Center
 Box 352145 Seattle, WA 98195
 COMMENT Contact: Daniel E. Geraghty (geraghty@fhcr.org)
 Overlapping Sequences:
 5': UWGC:370m23.013 (Genbank Accession: AC005530)
 3': UWGC:y67c112 (Genbank Accession: AC004211)
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.
 Double stranded (DS) coverage: 75.5%
 DS or two chemistry coverage: 98.9%
 Single stranded regions: 3
 Sequence Validation:
 This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BgIII		HindIII		NsiI	
Map	Seq	Map	Seq	Map	Seq
1069.11	1050.00	889.55	866.00	30541.40	30653.00
20320.67	20855.00	1050.18	1015.00	3279.08	3231.00
2171.50	2147.00	7268.78	7196.00		
2560.20	2531.00	10085.80	9992.00		
4335.42	4269.00	11212.78	11131.00		
2698.62	2628.00				
1927.50	1887.00				
3130.46	3090.00				
2166.69	2129.00				
2044.67	2005.00				

FEATURES
 source Location/Qualifiers
 1..47323

BASE COUNT 11556 a 11489 c 12284 g 11994 t
 ORIGIN

Query Match 47.8%; Score 109.4; DB 11; Length 47323;
 Best Local Similarity 95.0%; Pred. No. 2.7e-24;

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="6p21"
 /sub_clone="UMGC:370M23.002"
 3647..3932
 /rpt_family="Alu"
 complement(4999..5277)
 /rpt_family="Alu"
 6285..6572
 /rpt_family="Alu"
 complement(6972..7050)
 /rpt_family="MLT1"
 7286..7584
 /rpt_family="Alu"
 complement(8164..8609)
 /rpt_family="Alu"
 complement(21287..21895)
 /rpt_family="Alu"
 22715..22857
 /rpt_family="Alu"
 25510..25802
 /rpt_family="Alu"
 27835..28010
 /rpt_family="MER20"
 31295..31594
 /rpt_family="Alu"
 33515..33767
 /rpt_family="Alu"
 34223..34390
 /rpt_family="MIR"
 37372..37648
 /rpt_family="Alu"
 38526..38700
 /rpt_family="MER3"
 39583..40010
 /rpt_family="Alu"
 40046..40156
 /rpt_family="Alu"
 43194..43372
 /rpt_family="MERS"
 43325
 /note="clonal variation with 3' overlapping clone"
 44149
 /note="clonal variation with 3' overlapping clone"
 44451
 /note="clonal variation with 3' overlapping clone"
 44537
 /note="clonal variation with 3' overlapping clone"
 44814
 /note="clonal variation with 3' overlapping clone"
 44965
 /note="clonal variation with 3' overlapping clone"
 45760
 /note="clonal variation with 3' overlapping clone"
 45900
 /note="clonal variation with 3' overlapping clone"
 46851
 /note="clonal variation with 3' overlapping clone"
 46859
 /note="clonal variation with 3' overlapping clone"
 47032
 /note="clonal variation with 3' overlapping clone"
 47240..47256
 /note="clonal variation with 3' overlapping clone"
 /note="clonal variation with 3' overlapping clone -
 insertion of 17bp repeat"

```

Matches 113; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ACCGGACATTCAGTCTCTCCATCCAGGAGGCGAGTGGCCACTATGGGCTCTGGGCT 60
|||||
Db 35121 ACCGGACATTCAGTCTCTCCATCCAGGAGGCGAGTGGCCACTATGGGCTCTGGGCT 35180
|||||
QY 61 GCCCCTTGCTCTCTGACCTCTCTGGCAGCTCAGTGAACAGGCGGGGTATGA 119
|||||
Db 35181 GCCCCTTGCTCTCTGACCTCTCTGGCAGCTCAGTGAACAGGCGGGGTATGA 35239
|||||

RESULT 4
HSDJ144C9/c 77322 bp DNA PRI 22-NOV-1999
LOCUS Human DNA sequence from clone RPI-144C9 on chromosome 1p34.3-36.11,
DEFINITION complete sequence.
ACCESSION AL096774
VERSION AL096774.9 GI:6465842
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 77322)
AUTHORS Hall,R.
JOURNAL Direct Submission
Submitted (22-NOV-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Nov 22, 1999 this sequence version replaced gi:6138775.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
This sequence was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chrl
RPI-144C9 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pCYPAC2
This sequence is the entire insert of clone RPI-144C9.
FEATURES
Source Location/Qualifiers
1..77322
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="p34.3-36.11"
/clone="RPI-144C9"
/clone_11b="RPCI-1"
complement(1710..2227)
/note="match: GSS: Em:AQ355665"
complement(1780..2227)
/note="match: GSS: Em:AQ355618"
2858..3151
/note="match: GSS: Em:AQ627537.1"
complement(11439..11580)
/note="match: GSS: Em:AQ031787"
complement(24560..25030)

```

```

/note="match: GSS: Em:AQ07331"
complement(24637..25031)
/note="match: GSS: Em:AQ234890"
27999..28360
/note="match: STS: Em:G14651"
29391..29649
/note="match: STS: Em:G06203"
complement(37609..37922)
/note="match: GSS: Em:AQ114724"
47849..48201
/note="match: GSS: Em:AQ553574.1"
63243..63496
/note="match: GSS: Em:AQ142826"
BASE COUNT 21006 a 18901 c 18290 g 19125 t
ORIGIN

Query Match 15.0%; Score 34.4; DB 10; Length 77322;
Best Local Similarity 63.1%; Pred. No. 1.7;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 14 TGTCTCTCCATCCAGGAGGCGAGTGGCCACTATGGGCTCTGGGCTCTGGCTCTC 73
|||||
Db 31300 TGTCTCTCCATCCAGGAGGCGAGTGGCCACTATGGGCTCTGGGCTCTGGCTCTC 31241
|||||
QY 74 CTCCTGACCCCTCTGGCAGCTCA 97
|||||
Db 31240 CTCCTGACCCCTCTGGCAGCTCA 31217
|||||

RESULT 5
AC016716/c 211190 bp DNA HTG 04-DEC-1999
LOCUS Homo sapiens clone RP11-31213, *** SEQUENCING IN PROGRESS ***, 31
DEFINITION Unordered pieces.
ACCESSION AC016716
VERSION AC016716.1 GI:6524349
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 211190)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 211190)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H_NH0312103.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2127: contig of 2127 bp in length
* 2128 2145: gap of unknown length
* 4226 4226: contig of 2081 bp in length
* 4227 4244: gap of unknown length
* 4245 6447: contig of 2203 bp in length
* 6448 6465: gap of unknown length
* 6466 9092: contig of 2627 bp in length
* 9093 9110: gap of unknown length
* 9111 11395: contig of 2085 bp in length
* 11396 11213: gap of unknown length
* 11214 14033: contig of 2820 bp in length
* 14034 14051: gap of unknown length

```


RESULT	7
OSU25430	
LOCUS	OSU25430 1801 bp mRNA 23-MAY-1995
DEFINITION	Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete cds.
ACCESSION	U25430
VERSION	U25430.1 GI:818848
KEYWORDS	.
SOURCE	rice.
ORGANISM	Oryza sativa
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;

Db 148391 CAACACCCTGCAAGAAAGTCGCCCTCTCCACGCTTCTGGCATC 148434

RESULT	7						
OSU25430							
LOCUS	OSU25430	1801 bp	mRNA	PUN	23-MAY-1995		
DEFINITION	Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete cds.						
ACCESSION	U25430						
VERSION	U25430.1	GI:818848					
KEYWORDS							
SOURCE	rice.						
ORGANISM	Oryza sativa						
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
	euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;						

RESULT 9
AC008878
LOCUS AC008878 291288 bp DNA HTG 31-OCT-1999
DEFINITION Homo sapiens chromosome 19 clone CITB-HL_2207023, *** SEQUENCING IN
PROGRESS ***, 160 unordered pieces.
ACCESSION AC008878
VERSION AC008878.2 GI:6165135
KEYWORDS HTG; HTGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 291288)
TITLE Sequencing of Human Chromosome 19
AUTHORS Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 291288)
TITLE Direct Submission
AUTHORS DOE Joint Genome Institute.
JOURNAL
TITLE Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 31, 1999 this sequence version replaced gi:5686195.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 160 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1016: contig of 1016 bp in length
* 1017 gap of unknown length
* 1601: contig of 585 bp in length
* 1602 gap of unknown length
* 2427: contig of 826 bp in length
* 2428 gap of unknown length
* 3083: contig of 656 bp in length
* 3084 gap of unknown length
* 3959: contig of 876 bp in length
* 3960 gap of unknown length
* 7880: contig of 3921 bp in length
* 8576: gap of unknown length
* 8577: contig of 696 bp in length
* 12400: gap of unknown length
* 12401: contig of 3824 bp in length
* 13210: gap of unknown length
* 13211: contig of 810 bp in length
* 13929: gap of unknown length
* 13930: contig of 719 bp in length
* 16891: gap of unknown length
* 16892: contig of 2962 bp in length
* 17522: gap of unknown length
* 17523: contig of 631 bp in length
* 18054: gap of unknown length
* 17523: contig of 532 bp in length
* 18055: gap of unknown length
* 18736: contig of 682 bp in length
* 18737: gap of unknown length
* 19619: contig of 883 bp in length
* 19620: gap of unknown length
* 20466: contig of 847 bp in length
* 20467: gap of unknown length
* 21409: contig of 943 bp in length
* 21410: gap of unknown length
* 22300: contig of 891 bp in length
* 22301: gap of unknown length
* 23263: contig of 963 bp in length
* 23264: gap of unknown length
* 24270: contig of 1007 bp in length
* 24271: gap of unknown length
* 25014: contig of 744 bp in length
* 25015 gap of unknown length
* 25665: contig of 651 bp in length
* 27090: gap of unknown length
* 28065: contig of 1425 bp in length
* 28066: gap of unknown length
* 28837: contig of 975 bp in length
* 28837: gap of unknown length
* 28837: contig of 772 bp in length
* 28838: gap of unknown length
* 28838: contig of 3510 bp in length
* 32347: gap of unknown length
* 32348: contig of 771 bp in length
* 33118: gap of unknown length
* 33119: contig of 639 bp in length
* 33757: gap of unknown length
* 34621: contig of 864 bp in length
* 34622: gap of unknown length
* 35157: contig of 536 bp in length
* 35158: gap of unknown length
* 36196: contig of 1039 bp in length
* 36197: gap of unknown length
* 37333: contig of 1137 bp in length
* 37334: gap of unknown length
* 38054: contig of 721 bp in length
* 38055: gap of unknown length
* 38762: contig of 708 bp in length
* 38763: gap of unknown length
* 39524: contig of 762 bp in length
* 39525: gap of unknown length
* 40176: contig of 652 bp in length
* 40177: gap of unknown length
* 42851: contig of 2675 bp in length
* 42852: gap of unknown length
* 43760: contig of 909 bp in length
* 43761: gap of unknown length
* 44291: contig of 531 bp in length
* 44292: gap of unknown length
* 44840: contig of 549 bp in length
* 44841: gap of unknown length
* 45627: contig of 787 bp in length
* 45628: gap of unknown length
* 46298: contig of 671 bp in length
* 46299: gap of unknown length
* 47167: contig of 869 bp in length
* 47168: gap of unknown length
* 47878: contig of 711 bp in length
* 47879: gap of unknown length
* 48711: contig of 833 bp in length
* 48712: gap of unknown length
* 48959: contig of 248 bp in length
* 48960: gap of unknown length
* 50309: contig of 1350 bp in length
* 50310: gap of unknown length
* 51028: contig of 719 bp in length
* 51029: gap of unknown length
* 51872: contig of 844 bp in length
* 51873: gap of unknown length
* 52523: contig of 651 bp in length
* 52524: gap of unknown length
* 53417: contig of 894 bp in length
* 53418: gap of unknown length
* 54178: contig of 761 bp in length
* 54179: gap of unknown length
* 54928: contig of 750 bp in length
* 54929: gap of unknown length
* 55629: contig of 701 bp in length
* 55630: gap of unknown length
* 56741: contig of 1112 bp in length
* 56742: gap of unknown length
* 58251: contig of 1510 bp in length
* 58252: gap of unknown length
* 59103: contig of 852 bp in length
* 59103: gap of unknown length

```

RESULT 10
AC016298      62181 bp      DNA      HTG      24-NOV-1999
LOCUS      Homo sapiens clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING.
DEFINITION
AC016298
ACCESSION
VERSION
KEYWORDS
HTG; HTGS_PHASE0.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 62181)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-655M17
Unpublished
2 (bases 1 to 62181)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castelle,A., Colangelo,M., Collins,S., Collamore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McGowan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L4315
Center clone name: 655_M_17
-----
* NOTE: This record contains 78 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 775: contig of 775 bp in length
* gap of unknown length
* 776 1583: contig of 808 bp in length
* gap of unknown length
* 1584 2371: contig of 788 bp in length
* gap of unknown length
* 2372 3185: contig of 814 bp in length
* gap of unknown length
* 3186 3980: contig of 795 bp in length
* gap of unknown length
* 3981 4778: contig of 798 bp in length
* gap of unknown length
* 4779 5584: contig of 806 bp in length
* gap of unknown length
* 5585 6392: contig of 808 bp in length

```

QY	6	GACTCAGTGTCTCCATATCCGAGACCGCATGGCACATATGGGTCTGGGTGCCCC	65
Db	11905	GAITTTAAATTCTTCACCCTGANGTGAGTGTCTCTCTTTGGTGGGTGCCCC	11964
QY	66	TTCCTCTCTCTGAGCTCCCTGGCAGCTCAATGGACAAGGCCGGGTATGACTTGC	125
Db	11965	TCATCTTCCCTTGGCCCTGGGTGGCAGTCCCTGGANNAAGTCTCTACATCCCTTGC	12024
QY	126	AACCTGAAGCTGAA	138

Db 12025 GCCTGGGTGCA 12037

RESULT 11
HSJ773A18/c
LOCUS

DEFINITION HSJ773A18 128379 bp DNA PRI 23-NOV-1999
Human DNA sequence from clone 773A18 on chromosome 1p13.2-21.1
Contains genes for RAS-RELATED PROTEIN RAP-1A (C21KG) (KREV-1
PROTEIN) (GTP-BINDING PROTEIN SMG-P21A) (G-22K), KCND3 (potassium
voltage-gated channel, Shal-related subfamily, member 3), PROBABLE
ATP-DEPENDENT RNA HELICASE P47 HOMOLOG, ESTs, STSS, GSSs and CpG
islands, complete sequence.

ACCESSION
AL049557

VERSION 1

KEYWORDS HTG; C21KG; CPG Island; G-22K; GTP-BINDING PROTEIN; KCND3; KREV-1
PROTEIN; RAP1A; RAS-RELATED PROTEIN RAP1-A; SMG-P21A.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 128379)

Coville,G.

Direct Submission

REFERENCE Submitted (03-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Jul 27, 1999 this sequence version replaced gi:5579000.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.: EMBL; Sv.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information
on the WORMPEP database can be found at
<http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence is
the entire insert of clone 773A18. This sequence has been finished
according to sequence map criteria as follows. An attempt is made
to resolve all sequencing problems, such as compressions and
repeats, but not necessarily within known annotated human repeat
sequence elements (e.g. Alu). Where the sequence is ambiguous,
there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 1, constructed by the Sanger Centre Chromosome 1
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chrl>
773A18 is from the library RPCI-4 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR>: PCYPAC2.

FEATURES

source

1..128379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="p13.2-21.1"
/clone_lib="RPCI-4"
/clone="RP4-773A18"
complement(1..988)
/gene="dj773A18.3"
complement(1..988)
/gene="dj773A18.3"
/note="dj773A18.3 (similar to cytokeratin 18); match:
cDNAs: Em:X12876; match: proteins: Sw:P05784 Sw:P05783"
/codon_start=1
/pseudo

/evidence=not_experimental

misc_feature

1487..1866

/note="match: GSS: Em:AQ518605"

repeat_region

1535..1556

/note="11 copies 2 mer tt 100% conserved"

repeat_region

1557..1697

repeat_region
1872..1977
/note="MIR repeat: matches 76..206 of consensus"
repeat_region
1979..2236
/note="U6 repeat: matches 2..107 of consensus"
repeat_region
2237..2539
/note="L1MC1 repeat: matches 6071..6323 of consensus"
repeat_region
2540..2597
/note="AluY repeat: matches 4..299 of consensus"
repeat_region
2915..3217
/note="L1MC1 repeat: matches 6013..6071 of consensus"
repeat_region
3222..3352
/note="AluSX repeat: matches 1..303 of consensus"
repeat_region
3414..3702
/note="AluSc repeat: matches 1..131 of consensus"
repeat_region
3855..4028
/note="AluSX repeat: matches 13..301 of consensus"
repeat_region
4074..4274
/note="MER58C repeat: matches 29..84 of consensus"
repeat_region
4316..4840
/note="L1MB6 repeat: matches 5411..5607 of consensus"
repeat_region
4841..5150
/note="L1MB6 repeat: matches 5594..6136 of consensus"
repeat_region
5151..5192
/note="AluSp repeat: matches 1..309 of consensus"
repeat_region
5776..5852
/note="L1MB6 repeat: matches 6136..6175 of consensus"
repeat_region
5862..6106
/note="MIR repeat: matches 129..209 of consensus"
repeat_region
6123..6223
/note="MIR repeat: matches 5..252 of consensus"
repeat_region
6330..6483
/note="MIR repeat: matches 28..138 of consensus"
repeat_region
6666..6727
/note="FRAM repeat: matches 1..156 of consensus"
repeat_region
7100..7429
/note="31 copies 2 mer aa 77% conserved"
repeat_region
7838..8216
/note="AluYA5 repeat: matches 1..308 of consensus"
repeat_region
8268..8525
/note="L2 repeat: matches 1148..1537 of consensus"
repeat_region
8581..8904
/note="L1MB3 repeat: matches 5916..6183 of consensus"
repeat_region
8906..9254
/note="MER46B repeat: matches 2..236 of consensus"
repeat_region
9234..9548
/note="L1M4 repeat: matches 4791..5146 of consensus"
repeat_region
9550..9605
/note="L1M4 repeat: matches 2658..2976 of consensus"
repeat_region
9606..9702
/note="Tigger3(Golem) repeat: matches 2970..3028 of
consensus"
repeat_region
9730..9809
/note="L1M4 repeat: matches 4324..4430 of consensus"
repeat_region
9821..10068
/note="AluJ/FRAM repeat: matches 228..305 of consensus"
repeat_region
10057..10171
/note="Tigger3(Golem) repeat: matches 2277..2547 of
consensus"
repeat_region
10172..10563
/note="Tigger3(Golem) repeat: matches 1..111 of consensus"
repeat_region
10564..10876
/note="L1MEC repeat: matches 2272..2658 of consensus"
repeat_region
10877..10971
/note="MER44A repeat: matches 1..333 of consensus"
repeat_region
10972..11047
/note="L1MEC repeat: matches 2480..2272 of consensus"
repeat_region
11048..11442
/note="L1PA2 repeat: matches 6069..6144 of consensus"
repeat_region
11474..11935
/note="L1MEC repeat: matches 2112..2481 of consensus"
repeat_region
11938..12338
/note="L1MEC repeat: matches 1520..2023 of consensus"
repeat_region
11938..12338
/note="L2 repeat: matches 1958..2404 of consensus"


```
RESULT 13
AC014817/c
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014817
VERSION AC014817.1 GI:6436518
KEYWORDS HTG: HTGS_PHASE2.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 26065)
AUTHORS Adams M. and Venter J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210440 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..26065
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 7321 a 5960 c 5929 g 6855 t
ORIGIN

Query Match 14.7%; Score 33.6; DB 43; Length 26065;
Best Local Similarity 59.4%; Pred. NO. 2.8;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 1 ACCGGGACTTACGTCTCCTCCATCCAGGAGCGAGTGCCACTATGGGGTCTGGGCT 60
Db 1415 ACCAGGCCATGAATGCTTCTTCCCGAGCGCTCATGGAGCGCTTTTGGCTCCGCT 1356
Qy 61 GCCCTTCTCTCTCTTTCACCTCTCTTGGCAGCTC 96
Db 1355 GCGGTAAATCCGCCACCCCTCTTGGCAGCTC 1320

RESULT 14
AC017080/c
LOCUS
DEFINITION Homo sapiens clone Rp11-469G4, *** SEQUENCING IN PROGRESS ***, 34
unordered pieces.
ACCESSION AC017080
VERSION AC017080.1 GI:6554067
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106791)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 106791)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H_NH0469G04.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
```

```
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2077: contig of 2077 bp in length
2078: gap of unknown length
2098: contig of 4000 bp in length
6098: contig of unknown length
6117: gap of unknown length
6118: contig of 2689 bp in length
8807: gap of unknown length
8827: contig of 2729 bp in length
11556: contig of 4000 bp in length
11576: gap of unknown length
13852: contig of 2277 bp in length
13853: gap of unknown length
13873: contig of 2968 bp in length
16860: gap of unknown length
18661: contig of 2494 bp in length
19354: gap of unknown length
19375: contig of 2827 bp in length
22201: contig of 4000 bp in length
22219: gap of unknown length
26238: gap of unknown length
26239: contig of 2301 bp in length
28540: gap of unknown length
28559: contig of 3075 bp in length
31633: gap of unknown length
31634: contig of 4000 bp in length
35653: gap of unknown length
35672: contig of 4000 bp in length
39671: gap of unknown length
39690: gap of unknown length
42315: contig of 2625 bp in length
42316: gap of unknown length
42334: contig of 4000 bp in length
43335: gap of unknown length
43336: contig of 2337 bp in length
48690: gap of unknown length
48709: contig of 2438 bp in length
51147: contig of 4000 bp in length
51166: gap of unknown length
51167: contig of 2642 bp in length
53809: gap of unknown length
53828: contig of 2064 bp in length
55891: contig of 4000 bp in length
55892: gap of unknown length
55911: contig of 2134 bp in length
58045: gap of unknown length
58063: contig of 4000 bp in length
62064: gap of unknown length
62082: contig of 4000 bp in length
66083: contig of 2011 bp in length
66102: contig of 4000 bp in length
68113: gap of unknown length
68132: contig of 2726 bp in length
70857: gap of unknown length
70877: contig of 4000 bp in length
74877: gap of unknown length
74896: contig of 4000 bp in length
78896: gap of unknown length
78914: contig of 4000 bp in length
82915: gap of unknown length
82933: gap of unknown length
85817: contig of 2884 bp in length
85818: gap of unknown length
85837: contig of 4000 bp in length
89837: gap of unknown length
89855: contig of 3644 bp in length
93500: gap of unknown length
93518: contig of 4000 bp in length
97519: gap of unknown length
97537: contig of 2664 bp in length
100201: gap of unknown length
100220: contig of 2552 bp in length
102021: gap of unknown length
102772: contig of 4000 bp in length
102791: gap of unknown length
102792: contig of 4000 bp in length.
Location/Qualifiers
1..106791
/organism="Homo sapiens"
```


Mouse sex comb on
Human MP52 cDNA. M
Retroviral DNA bas
Human eyal gene co
Rat GlcAT-P cDNA.

Run on: March 20, 2000, 18:37:52 ; Search time 70.18 Seconds
(without alignments)

Title: US-09-092-296-2
perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC
Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:★

Word size :

Number of hits that pass the threshold : 623170

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query		DB	ID	Description	
		Match	Length				
1	218	95.2	439	1	V84366	Human stomach carc	
	2	34.4	15.0	2923	X03426	Human secreted pro	
	3	32	14.0	1128	1	V28846	Mouse coxsackievir
C	4	30.4	13.3	580	1	T02964	Cell-cycle regulat
	5	30.4	13.3	1301	1	V50430	Mouse coxsackievir
	6	29.8	13.0	3510	1	T18696	RAP-1 radiation pr
C	7	29.8	13.0	291	1	V86365	EST clone AR34. Ne
	8	29	12.7	2259	1	Q05106	Sequence encoding
	9	29	12.7	1110	1	N60409	Human apolipoprote
C	10	29	12.7	1110	1	N50450	Sequence encoding
	11	29	12.7	597	1	V69099	ApoB4L protease cD
	12	29	12.7	936	1	G69101	ApoB4L2 protease
C	13	29	12.7	597	1	T18068	Human ApoB4L2 cDN
	14	29	12.7	936	1	T18070	Human ApoB4L2 cDN
	15	29	12.7	1157	1	T65957	Human apolipoprote
C	16	28.8	12.6	1785	1	V76986	Human brain-specific
	17	28.4	12.4	2095	1	V77953	Human betai, 6'-N-ac
	18	28.4	12.4	2432	1	T57045	cDNA encoding glyc
C	19	28	12.2	683	1	G63794	Bovine trypsin gen
	20	28	12.2	701	1	V63795	Bovine trypsinogen
	21	28	12.2	110000	1	V30458_1	Continuation (2 of
C	22	28	12.2	110000	1	V30459_1	Continuation (2 of
	23	27.8	12.1	5359	1	V50396	Rat NMDA receptor
	24	27.6	12.1	3088	1	V05628	ADP ribosylation f
C	25	27.6	12.1	5102	1	V10261	Human COL4A6 gene.
	26	27.4	12.0	6030	1	V1907	Rattus norvegicus
	27	27.2	11.9	2122	1	T49318	Wnt-10b gene. Expr
C	28	27.2	11.9	5535	1	X21355	Human BAI1 gene. N
	29	27.2	11.9	2161	1	X03034	Human IL-1ra BAC c
	30	27.2	11.9	11901	1	X02998	Human IL-1ra BAC c
C	31	27	11.8	1554	1	U71243	Serine protease fo
	32	27	11.8	1294	1	T60587	Protein cognate of
	33	26.8	11.7	2703	1	O83695	New TGF-beta fami
C	34	26.8	11.7	2341	1	T31601	Cartilage-derived
	35	26.8	11.7	2703	1	T59729	Human bone morpho
	36	26.8	11.7	2703	1	T59405	DNA encoding huma
C	37	26.8	11.7	2703	1	T61412	cDNA encoding huma
	38	26.8	11.7	2703	1	T69693	cDNA encoding huma
	39	26.8	11.7	2703	1	T98191	cDNA for human MP5

RESULT	1
V84366	
ID	V84366 standard; cDNA to mRNA; 439 BP.
AC	V84366;
DT	30-MAR-1999 (first entry)
DE	Human stomach carcinoma CDNA clone HP10408.
KW	Transmembrane protein; HP10408; human; stomach cancer; ds.
KE	
OS	Homo sapiens.
HF	Location/Qualifiers
FT	75..311
CDS	/*tag= a
FT	/note="cDNA comprising the coding region (minus
FT	the stop codon) is claimed (claim 3)"
FT	

10-DEC-1998. 702445.
03-JUN-1998; JP-144948.
03-JUN-1997; JP-144948.
(PROT-) PROTEGENE INC.
(SAGA) SAGAMI CHEM RES CENTRE.
Kato S, Sekine S, Yamaguchi T;
WPI: 99-045730/04.
P-PSDB; W88498.
New human proteins containing transmembrane domains and their
encoding sequences - useful in the preparation of antibodies and
large-scale protein production, gene diagnosis, and gene therapy
Claim 4: Page 135; 178pp; English.
This is the nucleotide sequence of cDNA clone HPl0408, which
includes a coding region (also claimed) for a novel human
transmembrane protein (see W88498). The clone was isolated from a
stomach cancer cDNA library using a signal sequence detection
method, and by protein synthesis by in vitro translation. The
encoded protein has a putative signal sequence and a putative
internal transmembrane domain. The invention provides nucleotide
sequences (see W84359-76) coding for 18 transmembrane proteins
(see W88491-508), vectors containing such polynucleotides, and
eukaryotic cells containing the vectors. The proteins can be
used as antigens or as compositions in the preparation of
antibodies against the proteins. The polynucleotides can be used
as probes for gene diagnosis, and as gene sources for gene therapy
and large-scale production of proteins encoded by the cDNA. The
host cells are used for the detection of ligands corresponding to
the expressed proteins, and the screening of low mol.wt. medicines.
Sequence 439 Bp; 89 A; 137 C; 109 G; 10 T;

```

Query Match      95.2%; Score 218; DB 1; Length 439;
Best Local Similarity 99.6%; Pred. No. 2.5e-60;
Matches 229; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ACCGGGACTTCAGTGTCTCTCCATCCACGAGGCGAGTGGCCACTATGGGGTCTGGGCT 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 29 ACCGGGACTTCAGTGTCTCTCCATCCACGAGGCGAGTGGCCACTATGGGGTCTGGGCT 88
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 GCCCCTTGTCCTCCTTTGACCCCTCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 89 GCCCCTTGTCCTCCTCTTGACCCCTCTTGGCAGCTCACATGGAACAGGGCCGGGTATGAC 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 TTTCGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCCCTATGAGTCAGCTTCCCT 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 149 TTTCGCACTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCCCTATGAGTCAGCTTCCCT 208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 181 GGAATTGCTTGAAAAG-TCTGGCCTCTCCTCTCATCTCCCTTCTCAGGGACCA 229
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


<p> FH Key Location/Qualifiers cds 1..595 /*tag= a /label= ApoE4L protease </p> <p> PN W09413798-A. PD 23-JUN-1994. PF 16-DEC-1993; E03581. PR 16-DEC-1992; CA-085924. PR 04-MAR-1993; US-291401. PA (BERG/) BERGMANN J E. PA (BERG/) PREDDIE R E. PI Bergmann JE, Preddie RE; PI WPI: 94-234212/28. P-PSDB: R59841. DR P-PSDB: R59841. PT New proteinase esterase-like proteins - used to develop prods. PT for the diagnosis and treatment of Alzheimer's disease and PT related diseases CC Claim 9; Page 38-39; 72pp; English. CC The cDNA encodes ApoE4L, a protease catalyzing the formation of the CC abnormal beta/A4 variant of beta-amyloid protein, which is used CC to develop an inhibitor for the diagnosis and treatment of Alzheimer CC disease, Downs syndrome, Parkinson disease, schizophrenia, CC hyperlipoproteinemia or cardiovascular disease. CC Sequence 597 BP; 64 A; 240 C; 197 G; 96 T; </p>	<p> QY 6 GACTTCAGTGTCTCTCATCCAGAGCGAGTGCCACTATGGGTCTGGGTGCTGCCC 65 DB 305 GAGCGCGGCTCAGCGGCTCCCGAGCGCTGGGCGCCCTGCTGGACACAGCGCGGTG 246 </p> <p> QY 66 TTGTCTCTCTTGACCTCCTTTGGCAGCTCACAATGGAACAGGCGCGG 114 DB 245 CGGCGCGCACTGTGGGTCTCCTGGCGCGCCAGCGCGCTACAGGAGCGG 197 </p>
<p> RESULT 12 Q96101/C ID Q69101 standard; cDNA; 936 BP. AC Q69101; DT 26-JAN-1995 (first entry) DE ApoE4Lx2 protease cDNA. KW ApoE4Lx2; protease; enzyme; Alzheimer disease; diagnostic; KE therapeutic; ss. OS Homo sapiens. FH Key Location/Qualifiers cds 1..934 /*tag= a /label= ApoE4Lx2 protease </p> <p> PN W09413798-A. PD 23-JUN-1994. PF 16-DEC-1993; E03581. PR 16-DEC-1992; CA-085924. PR 04-MAR-1993; US-291401. PA (BERG/) BERGMANN J E. PA (BERG/) PREDDIE R E. PI Bergmann JE, Preddie RE; PI WPI: 94-234212/28. P-PSDB: R59843. DR P-PSDB: R59843. PT New proteinase esterase-like proteins - used to develop prods. PT for the diagnosis and treatment of Alzheimer's disease and PT related diseases CC Claim 13; Page 43; 72pp; English. CC The cDNA encodes ApoE4Lx2, a protease catalyzing the formation of the CC abnormal beta/A4 variant of beta-amyloid protein, which is used CC to develop an inhibitor for the diagnosis and treatment of Alzheimer CC disease, Downs syndrome, Parkinson disease, schizophrenia, CC hyperlipoproteinemia or cardiovascular disease. CC Sequence 936 BP; 149 A; 317 C; 336 G; 134 T; </p>	<p> QY 6 GACTTCAGTGTCTCTCATCCAGAGCGAGTGCCACTATGGGTCTGGGTGCTGCCC 65 DB 305 GAGCGCGGCTCAGCGGCTCCCGAGCGCTGGGCGCCCTGCTGGACACAGCGCGGTG 246 </p> <p> QY 66 TTGTCTCTCTTGACCTCCTTTGGCAGCTCACAATGGAACAGGCGCGG 114 DB 245 CGGCGCGCACTGTGGGTCTCCTGGCGCGCCAGCGCGCTACAGGAGCGG 197 </p>

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACATATGGGCTGGCTGCCGCC 65
 DB 305 GAGCGGGGCTCAGCGCCATCCCGAGCGCCCTGGGGCCCTGTGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGGACAGGCGCGG 114
 DB 245 CGGGCGGCGCACTGTGGGCTCCCTGGCGGCGCAGCCGCTACAGGAGCGG 197

RESULT 13
 T18068/c
 ID T18068 standard; cDNA; 597 BP.
 AC T18068;
 DT 22-MAY-1996 (first entry)
 DE Human ApoE4 cDNA.
 KW Alzheimer disease; beta-amyloid peptide precursor; APP;
 KW antisense; hyperlipoproteinaemia; cardiovascular disease;
 KW gene therapy; apolipoprotein; ApoE4; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 1..597
 FT /*tag- a
 FT /transl_except= pos:505..507;_aa:Ala
 FT /transl_except= pos:553..555;_aa:Ser
 PN CA2126787-A.
 PD 28-DEC-1995.
 PF 27-JUN-1994; 126787.
 PR 27-JUN-1994; CA-126787.
 PA (PRED/) PREDDIE R.E.
 PI Bergmann JE, Preddie RE;
 DR WPI; 96-151776/16.
 DR P-PSDB; R92113.
 PT Proteins and nucleic acids associated with Alzheimer's disease -
 PT used to develop products for diagnosis of Alzheimer's disease and
 PT related conditions, hyperlipoproteinaemia or cardiovascular disease
 PS Claim 9: Fig 3A; 67pp; English.
 CC 2 Open reading frames (T18068 and T18069) identified in the
 CC apolipoprotein E (ApoE) antisense sequence code for proteins
 CC apoE4L (R92113) and apoE4L1 (R92114), respectively. Both ORFs
 CC are transcribed from an upstream regulatory region (T18074).
 CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to
 CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of
 CC the significance of these and other sequences (see also T18065-77)
 CC in the aetiology of Alzheimer's disease (AD) provides means for
 CC diagnosing AD and related diseases, for the design of therapeutic
 CC reagents (e.g. ribozymes or antibodies) and potentially for gene
 CC therapy.
 SQ Sequence 597 BP; 63 A; 241 C; 196 G; 97 T;

Query Match 12.7%; Score 29; DB 1; Length 597;
 Best Local Similarity 54.1%; Pred. No. 3.4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACATATGGGCTGGCTGCCGCC 65
 DB 305 GAGCGGGGCTCAGCGCCATCCCGAGCGCCCTGGGGCCCTGTGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGGACAGGCGCGG 114
 DB 245 CGGGCGGCGCACTGTGGGCTCCCTGGCGGCGCAGCCGCTACAGGAGCGG 197

RESULT 14
 T18070/c
 ID T18070 standard; cDNA; 936 BP.
 AC T18070;
 DT 22-MAY-1996 (first entry).

DE Human ApoE4Lx2 cDNA.
 KW Alzheimer disease; beta-amyloid peptide precursor; APP;
 KW antisense; hyperlipoproteinaemia; cardiovascular disease;
 KW gene therapy; apolipoprotein; ApoE4Lx2; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 1..597
 FT /*tag- a
 FT /transl_except= pos:505..507;_aa:Ala
 FT /transl_except= pos:553..555;_aa:Ser
 PN CA2126787-A.
 PD 28-DEC-1995.
 PF 27-JUN-1994; 126787.
 PR 27-JUN-1994; CA-126787.
 PA (PRED/) PREDDIE R.E.
 PI Bergmann JE, Preddie RE;
 DR WPI; 96-151776/16.
 DR P-PSDB; R92113.
 PT Proteins and nucleic acids associated with Alzheimer's disease -
 PT used to develop products for diagnosis of Alzheimer's disease and
 PT related conditions, hyperlipoproteinaemia or cardiovascular disease
 PS Claim 13: Fig 6A; 67pp; English.
 CC 2 Open reading frames (T18068 and T18069) identified in the
 CC apolipoprotein E (ApoE) antisense sequence code for proteins
 CC apoE4L (R92113) and apoE4L1 (R92114), respectively. Both ORFs
 CC are transcribed from an upstream regulatory region (T18074).
 CC ApoE4 comprises 2 alleles, one of which causes E4L and E4L1 to
 CC combine to form a hybrid ORF, ApoE4Lx2 (T18070). Elucidation of
 CC the significance of these and other sequences (see also T18065-77)
 CC in the aetiology of Alzheimer's disease (AD) provides means for
 CC diagnosing AD and related diseases, for the design of therapeutic
 CC reagents (e.g. ribozymes or antibodies) and potentially for gene
 CC therapy.
 SQ Sequence 936 BP; 149 A; 317 C; 336 G; 134 T;

Query Match 12.7%; Score 29; DB 1; Length 936;
 Best Local Similarity 54.1%; Pred. No. 4;
 Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 6 GACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACATATGGGCTGGCTGCCGCC 65
 DB 305 GAGCGGGGCTCAGCGCCATCCCGAGCGCCCTGGGGCCCTGTGTGAACAGGCGCGGTG 246

OY 66 TTGTCTCTCTCTTGACCTCTTTGGCAGCTCACATGGACAGGCGCGG 114
 DB 245 CGGGCGGCGCACTGTGGGCTCCCTGGCGGCGCAGCCGCTACAGGAGCGG 197

RESULT 15
 T06957
 ID T06957 standard; cDNA to mRNA; 1157 BP.
 AC T06957;
 DT 19-JUN-1996 (first entry)
 DE Human apolipoprotein-E (ApoE) cDNA.
 KW Recombinant; human; apolipoprotein-E; ApoE; insect cells; larva;
 KW Manduca sexta; Autographa californica nuclear polyhedrosis virus;
 KW haemolymph; lipid complex; biologically active; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT misc_feature 62..119
 FT /*tag- a
 FT note= "misc_signal"
 FT cds 62..1015
 FT /*tag- b
 PN US5472858-A.
 PD 05-DEC-1995.
 PF 04-JUN-1991; 709949.
 PR 04-JUN-1991; US-709949.
 PA (WISC) WISCONSIN ALUMNI RES FOUND.
 PI Attie AD, Beckage NE, Gretsch DG, Sturley SL;
 DR WPI; 96-029812/03.
 DR P-PSDB; R86791.

5-6, 11 14
19-21
27-29 34 30-33
1-4, 12, 22-25
1-3, 1

10-18

Result No.	Score	Query			ID	Description
		Match	Length	DB		
C 1	30.4	13.3	580	4	US-08-627-610-7	Sequence 7, Appli
C 2	30.4	13.3	580	6	PCr-US95-04636-7	Sequence 7, Appli
C 3	29.8	13.0	3509	3	US-08-817-4368-1	Sequence 1, Appli
C 4	29.6	12.9	7218	1	US-08-233-463-14	Sequence 14, Appli
C 5	29	12.7	1157	1	US-07-709-949-1	Sequence 1, Appli
C 6	29	12.7	2259	7	5185254-3	Patent No. 5185252
C 7	28.4	12.4	2095	2	US-08-405-230-8	Sequence 8, Appli
C 8	28.4	12.4	2095	3	US-08-910-990-8	Sequence 8, Appli
C 9	27.6	12.1	5102	2	US-08-494-168-1	Sequence 1, Appli
C 10	27.6	12.1	3088	2	US-08-418-444-1	Sequence 1, Appli
C 11	27.2	11.9	2122	2	US-08-485-449-1	Sequence 1, Appli
C 12	27.2	11.9	3398	6	PCr-US95-08493-12	Sequence 12, Appli
C 13	27	11.8	1554	2	US-08-469-486-1	Sequence 1, Appli
C 14	27	11.8	1554	4	US-08-665-558-1	Sequence 1, Appli
C 15	27	11.8	1294	4	US-08-665-567-4	Sequence 4, Appli
C 16	26.8	11.7	3065	4	US-08-852-153-7	Sequence 7, Appli
C 17	26.8	11.7	2703	4	US-08-288-508-1	Sequence 1, Appli
C 18	26.6	11.6	1980	4	US-08-766-858-4	Sequence 4, Appli
C 19	26.6	11.6	1300	4	US-09-166-203-41	Sequence 41, Appli
C 20	26.4	11.5	3795	1	US-08-375-709-1	Sequence 1, Appli
C 21	26.4	11.5	8268	1	US-08-375-709-10	Sequence 10, Appli
C 22	26.4	11.5	3795	2	US-08-752-929-1	Sequence 1, Appli
C 23	26.4	11.5	8268	2	US-08-752-929-10	Sequence 10, Appli
C 24	26.4	11.5	1593	3	US-08-878-989-9	Sequence 9, Appli
C 25	26	11.4	7326	2	US-08-194-458-1	Sequence 1, Appli
C 26	26	11.4	44377	3	US-08-804-227-7	Sequence 7, Appli
C 27	26	11.4	44377	4	US-08-804-198-1	Sequence 1, Appli
C 28	25.8	11.3	2814	1	US-07-781-034-1	Sequence 1, Appli
C 29	25.8	11.3	2334	3	US-08-406-070-1	Sequence 1, Appli
C 30	25.8	11.3	655	3	US-08-812-845-2	Sequence 2, Appli
C 31	25.8	11.3	2814	6	PCr-US92-00382-1	Sequence 1, Appli
C 32	25.6	11.2	944	1	US-08-665-617-1	Sequence 1, Appli

QY 95 TCACATGGAACAGCGCGGTATGACTTTGCAACTG 130
Db 68 AGAAATGGTCTTCGCCGCCGTGAGATTGCTACAG 33

RESULT 2

PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/346,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..480
PCT-US95-04636-7

Query Match 13.3%; Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.25;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 35 GCAGTGGCCACTATGGGTCTGGGCTGCGCCCTTGCTCTCTTGACCTCTCTGGCAGC 94
Db 128 GCGGTGGCCAGCGCGCGGTGCTGCTGCGCCCAACATGCGCTTGTCCTGCTGCTGTC 69

QY 95 TCACATGGAACAGCGCGGTATGACTTTGCAACTG 130
Db 68 AGAAATGGTCTTCGCCGCCGTGAGATTGCTACAG 33

RESULT 3

US-08-817-436A-1
; Sequence 1, Application US/08817436A
; Patent No. 588280
; GENERAL INFORMATION:
; APPLICANT: Canaan1, Dan
; TITLE OF INVENTION: Human Checkpoint Gene and Gene for
; TITLE OF INVENTION: Antisense RNA thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wigman, Cohen, Leitner & Myers
; STREET: 900 17th Street, N.W., Suite 1000
; CITY: Washington

; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5 inch, 1.44mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,436A
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12445
; FILING DATE: 11-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/DOCKET NUMBER: 0744.066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-463-7700
; TELEFAX: 202-463-6915
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3509 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: ESTABLISHED XERODERMA PIGMENTOSUM GM2096-SV3
; IMMEDIATE SOURCE:
; CLONE: RAP-1 cdna
US-08-817-436A-1

Query Match 13.0%; Score 29.8; DB 3; Length 3509;
Best Local Similarity 51.9%; Pred. No. 0.91; Indels 0; Gaps 0;
Matches 67; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 11 CAGTGTCTCTCCATCCAGAGCGCAGTGGCGCTATGGGTCTGGGCTGCCCTTGTG 70
Db 1771 CACCGTCCCTCCATCCAGAGCGCAGAGAGAGATAACATCTCTATCTCTCTCTTGA 1830

QY 71 CTCTCTTGACCTCTTGGCAGCTCACATGGAACAGGCGCGGTATGACTTTGCAACTG 130
Db 1831 TACCTCTTGGAGCTTCCCAAGAAACAAAGAAAGAGAGAGATCTAGTTGCAGCTT 1890

QY 131 AAGCTGAAG 139
Db 1891 AAACGGAGG 1899

RESULT 4

US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299

Matches	80;	Conservative	0;	Mismatches	86;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Qy	63	CCCTGTGCTCCTCTTGACCCCTCTTGACAGCTC	CATGACAGCGCGGGTATGACTT	122
Db	487	CCATTCAGCAACAAGATTGTCACCTTCGACTTCC	CAACTTGCCCAATGCGCTGCAAAATGTTA	428
Qy	123	TGCAACTGAAGCTGAAGSGAGCTTTCTTGACA	AAATTCCTCTATGATGCCACGTCCTCTGG	182
Db	427	TCTAATAACACAGCAAGGTTTTCTTCAGATCA	TAAAGCTGTCTCATGCTCCAGCGATGGC	368
Qy	183	AATTGCTGAAAGTGTGCTCCTCCTCCATCTCC	CTTCCAGCGACC	228
Db	367	CCATCCACCACATTCCTGTTTCTTCGCGAGTGC	CTTGATGACC	322


```

RESULT          9
US-08-494-168-1/c
; Sequence 1, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Readers, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/494,168
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,465
; FILING DATE: 27-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28,665
; REFERENCE/DOCKET NUMBER: 40397/104/BABR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5102 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(2..82, 86..97, 101..4399, 4403..4420, 4424
; LOCATION: ..4465, 4469..4876, 4880..5101)
US-08-494-168-1

```

Query Match	12.1%	Score 27.6;	DB 2;	Length 5102;
Best Local Similarity	56.7%;	Pred. No. 6;		
Matches 51;	Conservative	0;	Mismatches 39;	Indels 0; Gaps 0;
QY	2	CCGGGACTTCAGTGTCTCCTCCATCCCAGGAGCGACGTGGCCCATGTGGGGTCTGGGCTG	61	
Dd	3992	CAGGGTCTCCGGTGTCGCCTTTGATTCAGAGTAGGCCACGGGGACCTGGGGGTCCACGGG	3933	
QY	62	CCCCCTGTCTCTCTTTGACCCTCTCTTGGC	91	

```

Db      3932  GGCCTGGCGGCGCTCGTCTCCATCTAGGC 3903
|||||  |||||  ||  |||||
RESULT  10
US-08-418-444A-1/c
; Sequence 1, Application US/08418444A
; Patent No. 5773698
; GENERAL INFORMATION:
; APPLICANT: KURODA, HISAO
; APPLICANT: HIROTA, NAOHICO
; APPLICANT: ITO, KAZUTOSHI
; TITLE OF INVENTION: GENE EXPRESSION VECTOR USING THE GENE
; TITLE OF INVENTION: GENE EXPRESSION REGULATING REGION OF THE ADP RIBOSYLATION
; TITLE OF INVENTION: FACTOR
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,444A
; FILING DATE: 07-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP HEI 6-71048
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 5773688man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 2589-024-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3088 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-418-444A-1

```

```

Query Match      12.1%; Score 27.6; DB 2; Length 3088;
Best Local Similarity 72.0%; Pred. No. 4.8;
Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      41  GCCACATATGGGTCTGGGCTGGCCCTTGTCTCTCTCTTGACCCCTCTTGG 90
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       64  GCCAGCAGGCGCCTGGCAGGCCCTTGGCCCCCTCTGCCCATCTTCGG 15

RESULT 11
US-08-485-449-1
; Sequence 1, Application US/08485449
; Patent No. 5824789
; GENERAL INFORMATION:
; APPLICANT: VANDENBERG, DAVID
; TITLE OF INVENTION: HEMATOPOIETIC GROWTH FACTORS, NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE ENCODING GROWTH FACTORS AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:

```


TELEPHONE: 617 542 5070
TELEFAX: 617 542 8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1554 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: CDS
LOCATION: 76..1551
US-08-469-486-1

Query Match 11.8%; Score 27; DB 2; Length 1554;
Best Local Similarity 48.4%; Pred. No. 5.6;
Matches 75; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 42 CCACATATGGGCTCTGGGCTGCCCCCTGCTCTCTGACCTCTTGGCAGCTCACATG 101
DB 71 CCACATATGGGCTCTGGGCTGCCCCCTGCTCTCTGACCTCTTGGCAGCTCACATG 130
QY 102 GAACAGGGCGGGTATGCACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCT 161
DB 131 GGCCGGGGGAGCGTGTCTGCTGCGCCGGGACGAGCCCGCTGCTGACAGAGGCC 190
QY 162 CCTATGAGTCCAGCTTCTCTGGAATGCTTGAAG 196
DB 191 GCAGGCCCACTCATTTCTTGAGGAGGTGAAGCAG 225

RESULT 14
US-08-469-658-1
Sequence 1, Application US/08469658
Patent No. 5917018
GENERAL INFORMATION:
APPLICANT: Th egeresen, Hans Christian
APPLICANT: Holtet, Thor Las
APPLICANT: Ezerodt, Michael
TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,658
FILING DATE: June 5, 1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/192,060
FILING DATE: February 4, 1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Paul T. Clark
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06363/002002

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617 542 5070
TELEFAX: 617 542 8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1554 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: YES
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: CDS
LOCATION: 76..1551
US-08-469-658-1

Query Match 11.8%; Score 27; DB 4; Length 1554;
Best Local Similarity 48.4%; Pred. No. 5.6;
Matches 75; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 42 CCACATATGGGCTCTGGGCTGCCCCCTGCTCTCTGACCTCTTGGCAGCTCACATG 101
DB 71 CCACATATGGGCTCTGGGCTGCCCCCTGCTCTCTGACCTCTTGGCAGCTCACATG 130
QY 102 GAACAGGGCGGGTATGCACTTTGCAACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCT 161
DB 131 GGCCGGGGGAGCGTGTCTGCTGCGCCGGGACGAGCCCGCTGCTGACAGAGGCC 190
QY 162 CCTATGAGTCCAGCTTCTCTGGAATGCTTGAAG 196
DB 191 GCAGGCCCACTCATTTCTTGAGGAGGTGAAGCAG 225

RESULT 15
US-08-665-647-4
Sequence 4, Application US/08665647
Patent No. 5935803
GENERAL INFORMATION:
APPLICANT: Dasquez, Nicki J.
APPLICANT: Ron, Dorit
APPLICANT: Voronova, Anna F.
APPLICANT: Napolitano, Eugene W.
TITLE OF INVENTION: METHODS TO IDENTIFY IMMUNOMODULATORS
TITLE OF INVENTION: USING COGNATE INTERACTION OF PKC-THETA
NUMBER OF SEQUENCES: 89
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW - Ste. 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,647
FILING DATE: 18-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H.
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 22550-20025.25
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 887-1500
TELEFAX: (202) 822-0168

; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1294 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
; US-08-665-647-4

Query Match 11.8%; Score 27; DB 4; Length 1294;
Best Local Similarity 46.5%; Pred. No. 5.1;
Matches 87; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 24 ATCCGAGGAGCGCAGTGGCCACTATGGGCTCTGGGCTGCCCCCTGTCTCTCTCTTGACCC 83
Db 578 ACCCAAGGATGGCAGGTGCGCTGCAGGACTGCATGTACAAAAGCCTGCAGAGGACT 637
QY 84 TCCTTGGCAGCTCACATGGAACAGGCGCGGTATGACTTTGCAACTGAAGCTGAAGGAGT 143
Db 638 GCGCGGCTACTCTGAGGGGACCGCAGCTGCTGAAGCGCATGCTCATCGGAGCTGT 697
QY 144 CTTTCTGACAAATTCCTCTATGAGTCCAGCTTCTCGAATTCCTTGAAGTCTGCCT 203
Db 698 GTGAGCCACAGATGCCACTACAGACTCCAGCCCGCCCGAGAGCATGGACGCTCTGCCT 757
QY 204 CCTCCTC 210
Db 758 CACCTC 764

Search completed: March 22, 2000, 01:22:33
Job time: 5637 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2000, 05:09:43 ; Search time 802.03 Seconds
(without alignments)
1078.048 Million cell updates/sec

Title: US-09-092-296-2
Perfect score: 229
Sequence: 1 ACCGGGACTTCAGTGTCTCC.....CCATCTCCTTCAGGGACCA 229

Scoring table: IDENTITY_NUC

Searched: 4538634 seqs, 1887831982 residues

Database : EST:*

Word size : 0

Number of hits that pass the threshold : 9077268

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*

- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_gss1:*
- 80: gb_gss2:*
- 81: gb_gss3:*
- 82: gb_gss4:*
- 83: em_gss1:*
- 84: em_gss2:*
- 85: em_gss3:*
- 86: em_gss4:*
- 87: gb_gss5:*
- 88: gb_gss6:*
- 89: gb_gss7:*
- 90: gb_gss8:*
- 91: gb_gss9:*
- 92: em_gss5:*
- 93: em_gss6:*
- 94: em_gss7:*
- 95: em_gss8:*
- 96: em_gss9:*
- 97: em_gss10:*
- 98: em_gss11:*
- 99: gb_gss10:*
- 100: gb_gss11:*
- 101: em_gss12:*
- 102: gb_gss12:*
- 103: gb_gss13:*
- 104: gb_gss14:*
- 105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
c 1	113	49.3	404	61	AI857998	AI857998 wj69b01.x
c 2	75.4	32.9	552	82	AQ718761	AQ718761 HS.5511.B
c 3	61.8	27.0	328	42	AI136523	AI136523 UI-R-C2p-
c 4	35	15.3	348	21	F06958	F06958 HSCIQC101 n
5	35	15.3	509	88	AQ838514	AQ838514 HS.5011.A
6	34.4	15.0	238	32	AA376266	AA376266 EST8915
c 7	34.4	15.0	412	38	AA769782	AA769782 an71b05.5

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (<http://www.hsc.washington.edu>)
Plate: 1087 row: L column: 18
Seq primer: T7
Class: BAC ends
High quality sequence stop: 552.

Location/Qualifiers
1. 552
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate-1087 Col-18 Row-L"
/clone_lib="RPC1-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"

FEATURES

source

BASE COUNT 141 a 153 c 124 g 117 t 17 others
ORIGIN

Query Match 32.9%; Score 75.4; DB 82; Length 552;
Best Local Similarity 87.2%; Pred. No. 3.4e-12;
Matches 82; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 103 AACAGGGCGGATGATCTTGAAGTGAAGTGAAGTCTTTCTGACAAATCTTC 162
Db 474 AGCATGGAGGTATGACTTTGCANCTGAAGTGAAGTCTTTTACAAATCTTC 415
QY 163 CTATGATCCAGCTCTCTGGAATCTTGAAGAAAG 196
Db 414 TTATGATCCAGCTCTCTGGAATCTTGAAGAAAG 381

RESULT 3

LOCUS A1136523/c

DEFINITION UI-R-C2p-ng-e-02-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
UI-R-C2p-ng-e-02-0-UI 3', mRNA sequence.

ACCESSION A1136523

VERSION A1136523.1

KEYWORDS GI:3637300

SOURCE EST.

ORGANISM Norway rat.

Rattus norvegicus.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 328)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

9704477

On Jan 19, 1998 this sequence version replaced gi:2150222.

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.wesg.uiowa.edu

The sequence tag present in the cDNA between the NotI site and the

oligo-dt track served to identify it as a clone from the normalized

adult Lung library. CDNA Library Preparation: M. Fatima Bonaldo,

Ph.D. Clone distribution: clones will be available through Research

Genetics

Seq primer: M13 Forward.

Location/Qualifiers

1. .328

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-C2p-ng-e-02-0-UI"

/clone_lib="UI-R-C2p"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

FEATURES

source

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p
library is a subtracted library derived from the UI-R-C1
library which is a subtracted library derived from the
UI-R-C0 library. The UI-R-C0 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
embryo. The tag is a string of 3-5 nucleotides present
between the Not I site and the oligo-dt track which allows
identification of the library of origin of a clone within
the mixture. The subtracted library (UI-R-C2p) was
constructed as follows: PCR amplified cDNA inserts from
UI-R-C1 clones from which 3' ESTs had been derived was
used as a driver in a hybridization with the UI-R-C1
library in the form of single-stranded circles. The
remaining single-stranded circles (subtracted library) was
purified by hydroxyapatite column chromatography,
converted to double-stranded circles and electroporated
into DH10B bacteria (Life Technologies) to generate the
UI-R-C2p library. This procedure has been previously
described (Bonaldo, Lennon and Soares, Genome Research 6:
791-806, 1996)

BASE COUNT 62 a 77 c 98 g 91 t
ORIGIN

Query Match 27.0%; Score 61.8; DB 42; Length 328;
Best Local Similarity 69.5%; Pred. No. 2.8e-08;
Matches 98; Conservative 0; Mismatches 42; Indels 1; Gaps 1;

QY 90 GCAGCTCAGTGAACAGCGCGGTATGACTTTGCAAGCTGAAGAGTCTTTTC 149
Db 327 GCAGCTCAGTGAACAGCGCGGTATGACTTTGCAAGCTGAAGAGTCTTTTC 268
QY 150 TCACAAATCTCTCTATGATCCAGCTTCCTGGAATGCTTGAAGAGTCTTCTCC 208
Db 267 AGGCCAGACCTCCCAAGACTCCGGCTTCTGACATGCTCCAAAGATCTGCTCTCC 208
QY 209 TCCATCTCTCTTTCAGGACCA 229
Db 207 TCCACCTCTCAGCGGGACCA 187

RESULT 4

LOCUS F06958

DEFINITION HSC1QC101 normalized infant brain cDNA Homo sapiens CDNA clone

C-1QC10, mRNA sequence.

ACCESSION F06958

VERSION F06958.1

KEYWORDS GI:672595

SOURCE EST.

human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 348)

Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,

Devignes, M.D., Duprat, S., Houllatte, R., Jumeau, M.N., Lamy, B.,

Lorenzo, F., Mitchell, R., Mariage-Samson, R., Pietu, G., Pouliot, Y.,

Sebastiani-Kabakthis, C. and Tessier, A.

IMAGE: molecular integration of the analysis of the human genome

and its expression

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

95277534

On Sep 21, 1992 this sequence version replaced gi:278976.

Contact: Genethon

GenexPress-Genethon

Genethon Centre de recherche sur le Genome Humain

1, rue de l'Internationale, BP50 91002 EVRY Cedex, FRANCE

Tel: 33169472800

Fax: 33160778698

Email: genexpress@genethon.fr

TITLE

JOURNAL

MEDLINE

COMMENT

The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavet@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 source
 1..238
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):180944"
 /db_xref="taxon:9606"
 /clone_lib="HSC172 cells II"
 /cell_type="fibroblast"
 /cell_line="HSC172 (60PDL)"
 /dev_stage="fetal"
 /note="Organ: lung; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"
 BASE COUNT 48 a 70 c 50 g 66 t 4 others
 ORIGIN

Query Match 15.0%; Score 34.4; DB 32; Length 238;
 Best Local Similarity 63.1%; Pred No. 3;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 14 TGTCTCTCCATCCAGGAGCGGCGGCTGCTGGGCTGCGCCCTTGTCTCTC 73
 ||| ||||| ||||| ||| ||||| || ||||| || ||||| || ||||| ||
 Db 141 TGTGCTCTCCCTCCAGGATCCCTTTGGTGAGTATGTTGTCAGGATGCACCACCAC 200
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 QY 74 CTCTGACCTCCCTTGGCAGCTCA 97
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 201 CTCTAGATACCTTCAGGCAACACA 224

RESULT 7
 AA769782/c
 LOCUS
 DEFINITION ah71b05.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321041 3',
 mRNA sequence.
 VERSION AA769782
 KEYWORDS AA769782.1 GI:2821020
 SOURCE EST
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 412)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 9, 1998 this sequence version replaced gi:936203.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 691 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 402.
 Location/Qualifiers
 1..412
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

FEATURES
 source
 1..412
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

/clone="1321041"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 111 a 78 c 116 g 107 t
 ORIGIN

Query Match 15.0%; Score 34.4; DB 38; Length 412;
 Best Local Similarity 63.1%; Pred. No. 3.7;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 14 TGTCTCTCCATCCAGGAGCGGCGGCTGCTGGGCTGCGCCCTTGTCTCTC 73
 ||| ||||| ||||| ||| ||||| || ||||| || ||||| || ||||| ||
 Db 328 TGTGCTCTCCCTCCAGGATCCCTTTGGTGAGTATGTTGTCAGGATGCACCACCAC 269
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 QY 74 CTCTGACCTCCCTTGGCAGCTCA 97
 ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
 Db 268 CTCTAGATACCTTCAGGCAACACA 245

RESULT 8
 AI632159/c
 LOCUS
 DEFINITION ts85el2.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:2238094 3',
 mRNA sequence.
 VERSION AI632159
 KEYWORDS AI632159.1 GI:4683489
 SOURCE EST
 ORGANISM human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 458)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 7, 1998 this sequence version replaced gi:3121420.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 439.
 Location/Qualifiers
 1..458
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2238094"
 /clone_lib="NCI-CGAP_GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"

FEATURES
 source

27 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or

/note="Vector: pSPORt1; Site_1: NotI; Site_2: SalI; 1st strand cDNA was primed with a Not I - oligo(dt)15 primer [5'] pGATCTAGTACGGCCAGCGGCCGCCCCCTTTTTCATT3'; double-stranded cDNA was ligated to Sal I adaptors (BRL), digested with Not I and cloned into the Not I and Sal I sites of the pSPORt1 vector (BRL). Library was constructed by Matthew Clark (Lehrsch lab: ICfG, London and Max planck Institut fuer Molekulare Genetik, Berlin). cDNAs for ESR analysis were selected following oligonucleotide hybridization fingerprinting of arrayed clones from zebrafish late somitogenesis (26 ss), adult liver or embryonic shield stage (5.6 h) libraries. Fingerprint data were used to computationally cluster cDNAs, and a single cDNA from each cluster was chosen for sequencing. In some cases multiple members of the same cluster were sequenced to assess clustering parameters or single clones were sequenced additional times to assess quality control.

CONCLOT.		
BASE COUNT	175 a	133 c
ORIGIN	112 g	103 t

```

Query Match      14.18; Score 32.4; DB 69; Length 523;
Best Local Similarity 56.68; pred No 16.

```

[illegible]

QY 92 AGCTCACATGGAACAGGGCGGGGTATGACTTGCACACTGAAGCTGA 137
 Db 398 ATCTGTGAGAAACGAGTGTCTGGATCTGGCTTGACTGGTTACTCTTA 353

RESULT	12
F08745	
LOCUS	F08745 339 bp mRNA
DEFINITION	HSC1DB01 normalized infant brain cDNA Homo sapiens cDNA clone c-1db01, mRNA sequence.
	EST 20-FEB-1995

ACCESSION F08745
VERSION F08745.1 GI:672165
KEYWORDS EST.
SOURCE human

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	

REFERENCE	AUTHORS
1 (pages 1 to 339)	
	Auffray, C., Behar, G., Bols, F., Boucher, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M., Lamy, B., Lorenzo, F., Mitchell, H., Marriage-Samson, R., Pletu, G., Pouliot, Y., Sorensen, K., Tardieu, A., Tardieu, C., Tardieu, D., Tardieu, E., Tardieu, F., Tardieu, G., Tardieu, H., Tardieu, I., Tardieu, J., Tardieu, K., Tardieu, L., Tardieu, M., Tardieu, N., Tardieu, O., Tardieu, P., Tardieu, Q., Tardieu, R., Tardieu, S., Tardieu, T., Tardieu, U., Tardieu, V., Tardieu, W., Tardieu, X., Tardieu, Y., Tardieu, Z.

TITLE	IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL	C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE	95277534
COMMENT	On 05-23-1997 10:02:43Z

ou sep 24, 1992 unis sequence version replaced gi:2729421
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE

Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.

```

FEATURES             source
1..339
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-ldb01"
Genexpress_library_id: C; Genexpress_sequence_id: y3c-ldb01
Seq primer: (-21)M13 universal
High quality sequence stop: 277.
Location/Qualifiers

```

[illegible]

ACCESSION AI351147
VERSION AI351147.1 GI:4088353
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 477)
REFERENCE NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.lln.gov/bbrp/image/image.html

Trace considered overall poor quality
Insert Length: 748 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. 477
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1947312"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"

FEATURES
source

oligo(dT) primer Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 75 a 150 c 174 g 78 t

ORIGIN

Query Match 14.18; Score 32.2; DB 45; Length 477;

Best Local Similarity 56.0%; Pred. No. 17;

Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Db 146 GAGCGCGTCTCAGGCCATCCGTGAGCCCTGGGGCCCTGATGGACAGGGCCCGTG 205

Qy 66 TTGTCTCTCTTGTGACCCCTCTGTGACGCTCACATGGAACAGGCGCGG 114
|||||
Db 206 CGGTCCGCACTGTGGCTCCCTGGCGGTGAGCGGTACAGGAGCGG 254

RESULT 15

AW009894 360 bp mRNA EST 10-SEP-1999
LOCUS ws8h09.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2505089 3',
DEFINITION mRNA sequence.
ACCESSION AW009894
VERSION AW009894.1 GI:5858672
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948428.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -400P from Gibco.

FEATURES

Location/Qualifiers
1..360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2505089"
/clone_lib="NCI_CGAP_Co3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library went through one round of
normalization."
BASE COUNT 108 a 66 c 78 g 108 t

ORIGIN

Query Match 14.08; Score 32; DB 63; Length 360;
Best Local Similarity 47.9%; Pred. No. 18;
Matches 92; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 24 ATCCAGGAGCGCAGTGGGCTCTGGGCTGCCCTTGTCTCTCTCTGACCC 83
|||
Db 57 ATACAAGAGTTGAGGTGTCGAAGCTTTAGCTTAAGCTCTATCTGTTATTAGGATTGAATC 116
|||
Qy 84 TCCTTGCACCTCACATGGAACAGCGCGGTATGACTTTCACACTGAGCTGAGGAGT 143
|||
Db 117 TCTTGAGTGGGCTGTGTAAGTATACCTCTAGGACTCAACTTGAAGAAGA 176
|||
Qy 144 CTTTCTGACAAATTCCTCTATGAGTCCAGCTTCTCGAAATGCTTGAAGAGTCTGCCT 203
|||
Db 177 ATTTGAGGACACCCGCGCACATGCAGATGGGACCAAGAGTGGCTGGAACATGATGCA 236
|||

Qy 204 CCTCTCTCACT 215
|||
Db 237 ACTCTACCATTT 248

Search completed: March 22, 2000, 05:09:44
Job time: 1365 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 21, 2000, 13:39:45 ; Search time 479.47 Seconds
(without alignments)
-1513.537 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GGCACCGGAGCTTCAGTGT.....CCCTTCAGGACCAAGCGTCA 239

Scoring table: IDENTITY_NUC

Searched: 821193 seqs, -1518192014 residues

Database : GenEmbl.*

Word size : 0

Number of hits that pass the threshold : 1642386

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_em.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_ov.*
- 23: em_ro.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_ro.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_un.*
- 31: em_vi.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*
- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*

50: gb_pl3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	113.4	47.4	192650	10	AB023048	AB023048 Homo sapi
2	113.4	47.4	200000	10	AP000511	AP000511 Homo sapi
3	113.4	47.4	47323	11	AC005937	AC005937 Homo sapi
4	34.6	14.5	291288	42	AC008878	AC008878 Homo sapi
5	34.4	14.4	77322	10	HS0144C9	AL096774 Human DNA
6	34.2	14.3	200000	41	AC006445	AC006445 Homo sapi
7	34	14.2	36676	11	HSJ858B16	AL096768 Human DNA
8	33.8	14.1	62181	44	AC016298	AC016298 Homo sapi
9	33.6	14.1	128379	11	HSJ773A18	AL049557 Human DNA
10	33.6	14.1	26065	43	AC014817	AC014817 Drosophil
11	33.6	14.1	106791	45	AC017080	AC017080 Homo sapi
12	33.4	14.0	211190	44	AC016716	AC016716 Homo sapi
13	33.4	14.0	191010	45	AC016764	AC016764 Homo sapi
14	33	13.8	1801	7	OSU23430	U25430 Oryza sativ
15	33	13.8	97580	11	HS591N18	AL031594 Human DNA
16	32.6	13.6	5280	9	AB011114	AB011114 Homo sapi
17	32.4	13.6	179532	42	AC010856	AC010856 Homo sapi
18	32.2	13.5	108049	43	AC012294	AC012294 Mus muscu
19	31.6	13.2	2803	2	TNU93354	U93354 Thermotoga
20	31.6	13.2	103150	8	ATF11111	AL079347 Arabidops
21	31.6	13.2	125502	8	ATF1420	AL023094 Arabidops
22	31.6	13.2	39503	11	AC005595	AC005595 Homo sapi
23	31.6	13.2	727	34	MITM117	L31623 Mitrocoma c
24	31.6	13.2	168986	40	AC008013	AC008013 Homo sapi
25	31.6	13.2	179622	43	AC009533	AC009533 Homo sapi
26	31.6	13.2	208079	44	AC013391	AC013391 Homo sapi
27	31.4	13.1	39261	10	AC000387	AC000387 Genomic s
28	31.4	13.1	100635	11	AC003104	AC003104 Homo sapi
29	31.4	13.1	189146	32	AP000846	AP000846 Homo sapi
30	31.4	13.1	303606	33	AL133353	AL133353 Homo sapi
31	31.4	13.1	183725	41	AC009558	AC009558 Homo sapi
32	31.4	13.1	154791	45	AC015478	AC015478 Homo sapi
33	31.2	13.1	177241	11	HS402G11	AL022328 Human DNA
34	31.2	13.1	78054	35	AC004296	AC004296 Drosophil
35	31.2	13.1	245767	42	AC011607	AC011607 Homo sapi
36	31.2	13.1	207043	44	AC011612	AC011612 Homo sapi
37	31.2	13.1	203734	44	AC011811	AC011811 Homo sapi
38	31	13.0	3509	5	AB0086	AB0086 Sequence 1
39	31	13.0	3510	9	HSX9050	X99050 H.sapiens m
40	31	13.0	42244	11	AC004660	AC004660 Homo sapi
41	31	13.0	97585	11	HS63G5	Z94160 Homo sapien
42	31	13.0	186448	32	AP000785	AP000785 Homo sapi
43	31	13.0	172461	33	AC006247	AC006247 Drosophil
44	31	13.0	195909	40	AF165124	AF165124 Homo sapi
45	31	13.0	120980	41	AC008611	AC008611 Homo sapi

ALIGNMENTS

RESULT	1
AB023048	
LOCUS	AB023048 192650 bp DNA PRI 20-NOV-1999
DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, clone:53L9, complete sequence.
ACCESSION	AB023048
VERSION	AB023048.1 GI:5672603
KEYWORDS	HTC.
SOURCE	Homo sapiens cell_line:978SK DNA, clone:53L9.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (sites)
Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

REFERENCE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

AUTHORS

Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

REFERENCE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

AUTHORS

Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

REFERENCE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

AUTHORS

Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

REFERENCE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

AUTHORS

Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

REFERENCE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

AUTHORS

Shina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,
Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,
Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,
Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,
Babram,S., and Inoko,H.
Molecular dynamics of MHC genesis unraveled by sequence analysis of
The 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
2007539

TITLE

2 (bases 1 to 192650)
The 1,796,938-bp HLA class I region

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)

MEDLINE

2007539

FEATURES

source

1. 200000

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/map="6p21.3"

108774..108864

/standard_name="D6S2088"

/note="SHGC-12985:The location is between each flanking site of PCR primers."

/db_xref="GDB:735268"

115648..115906

/standard_name="D6S1898"

/note="WI-9418:The location is between each flanking site of PCR primers."

/db_xref="GDB:678272"

complement(115665..115896)

/note="Cda0vhl0:The location is between each flanking site of PCR primers."

/db_xref="GDB:443382"

142030..142159

/standard_name="D6S2203"

/note="SHGC-16870:The location is between each flanking site of PCR primers."

/db_xref="GDB:741215"

complement(142085..142322)

/note="RH18132:The location is between each flanking site of PCR primers."

/db_xref="GDB:4573021"

complement(142479..142740)

/standard_name="D6S1851"

/note="SHGC-10808:The location is between each flanking site of PCR primers."

/db_xref="GDB:675281"

complement(164525..164665)

/note="SHGC-3064:The location is between each flanking site of PCR primers."

/db_xref="GDB:1234116"

complement(169209..169574)

/standard_name="D6S953"

/note="UT5233:The location is between each flanking site of PCR primers."

/db_xref="GDB:313481"

BASE COUNT 52605 a 47531 c 49366 g 50498 t

ORIGIN

Query Match 47.4%; Score 113.4; DB 10; Length 200000;

Best Local Similarity 95.1%; Pred. No. 2e-25;

Matches 117; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GGCCACGGGACTTCAGTGTCTCTCCATCCAGGAGCGGCGGCTA 120

Db 109958 GGCCACGGGACTTCAGTGTCTCTCCATCCAGGAGCGGCGGCTA 120

Qy 61 GGCTGCCCTTGTCTCTCTTGCAGCTCCTTGCAGCTCAGTGAACAGGCGGCTA 120

Db 110018 GGCTGCCCTTGTCTCTCTTGCAGCTCCTTGCAGCTCAGTGAACAGGCGGCTA 120

Qy 121 TGA 123

Db 110078 AGA 110080

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES	Source	Location/Qualifiers
AC005937	47323 bp	DNA
LOCUS	05-NOV-1998	
DEFINITION	Homo sapiens clone UWGC:370M23.002 from 6p21, complete sequence.	
ACCESSION	AC005937	
VERSION	AC005937.1	GI:3845393
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 47323)	
TITLE	Janer, M., Guillaudoux, T., Vu, Q., Kutayavin, T., Harter, H. and Geraghty, D.E.	
JOURNAL	Large scale sequence analysis of the human MHC class I region	
REMARK	Unpublished (1998)	
	Fred Hutchinson Cancer Research Center	
	The Clinical Research Division	
	1100 Fairview Ave. N., P.O. Box 19024	
	Seattle, WA 98109-1024	
	2 (bases 1 to 47323)	
	Geraghty, D.E. and Olson, M.V.	
	Direct Submission	
	Submitted (05-NOV-1998) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA	
	University of Washington Human Genome Center	
	Box 352145 Seattle, WA 98195	
	Contact: Daniel E. Geraghty (geraghty@fhcr.org)	
COMMENT	Overlapping Sequences:	
	5': UWGC:370M23.013 (Genbank Accession: AC005530)	
	3': UWGC:367C112 (Genbank Accession: AC004211)	
	Sequence Quality Assessment:	
	This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.	
	All manually edited bases have been reduced to quality zero.	
	Quality levels above 40 are expected to have less than 1 error in 10,000 bp.	
	Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.	
	Double stranded (DS) coverage: 75.5%	
	DS or two chemistry coverage: 98.9%	
	Single stranded regions: 3	
	Sequence Validation:	
	This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below.	
	Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.	
	BglII	HindIII
	Map	Seq
	Map	Seq
	Map	Seq
	1069.11	1050.00
	889.55	866.00
	30541.40	30653.00
	20320.67	20855.00
	1050.18	1015.00
	3279.08	3231.00
	2171.50	2147.00
	7268.78	7196.00
	2560.20	2531.00
	10085.80	9992.00
	4335.42	4269.00
	11212.78	11131.00
	2698.62	2628.00
	1927.50	1887.00
	47240.47	47256.00

```
/note="clonal variation with 3' overlapping clone -
insertion of 17bp repeat"
BASE COUNT 11556 a 11489 c 12284 g 11994 t
ORIGIN

Query Match 47.4%; Score 113.4; DB 11; Length 47323;
Best Local Similarity 95.1%; Pred. No. 1.9e-25;
Matches 117; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCCACCGGACATTCCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGGTCTG 60
Db 35117 GCCACCGGACATTCCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGGTCTG 35176

QY 61 GGCTGCCCTTGCTCCCTCTTGTGACCTCTTGGCAGCTCACATGGAACAGGCGCGGTA 120
Db 35177 GGCTGCCCTTGCTCCCTCTTGTGACCTCTTGGCAGCTCACATGGAACAGGCGCGGCT 35236

QY 121 TGA 123
Db 35237 AGA 35239

RESULT 4
AC008878 291288 bp DNA HTG 31-OCT-1999
LOCUS Homo sapiens chromosome 19 clone C19B-H1_2207023, *** SEQUENCING IN
DEFINITION AC008878
ACCESSION AC008878
VERSION AC008878.2 GI:6165135
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 291288)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 19
REFERENCE JOURNAL
Unpublished
2 (bases 1 to 291288)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 4800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Oct 31, 1999 this sequence version replaced gi:5686195.
www.jgi.doe.gov.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 160 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1016: contig of 1016 bp in length
* gap of unknown length
* 1017 1601: contig of 585 bp in length
* gap of unknown length
* 1602 2427: contig of 826 bp in length
* gap of unknown length
* 2428 3083: contig of 656 bp in length
* gap of unknown length
* 3084 3959: contig of 876 bp in length
* gap of unknown length
* 3960 7880: contig of 3921 bp in length
* gap of unknown length
* 7881 8576: contig of 696 bp in length
* gap of unknown length
* 8577 12400: contig of 3824 bp in length
* gap of unknown length
* 12401 13210: contig of 810 bp in length
* gap of unknown length
* 13211 13929: contig of 719 bp in length
* gap of unknown length

16891: gap of unknown length
16891: contig of 2962 bp in length
17522: gap of unknown length
17522: contig of 631 bp in length
18054: gap of unknown length
18054: contig of 532 bp in length
18736: gap of unknown length
18736: contig of 682 bp in length
19619: gap of unknown length
19619: contig of 883 bp in length
20466: gap of unknown length
20466: contig of 847 bp in length
21409: gap of unknown length
21409: contig of 943 bp in length
22300: gap of unknown length
22300: contig of 891 bp in length
23263: gap of unknown length
23263: contig of 963 bp in length
24270: gap of unknown length
24270: contig of 1007 bp in length
25014: gap of unknown length
25014: contig of 744 bp in length
25665: gap of unknown length
25665: contig of 651 bp in length
27090: gap of unknown length
27090: contig of 1425 bp in length
28065: gap of unknown length
28065: contig of 975 bp in length
28837: gap of unknown length
28837: contig of 772 bp in length
32347: gap of unknown length
32347: contig of 3510 bp in length
33118: gap of unknown length
33118: contig of 771 bp in length
33757: gap of unknown length
33757: contig of 639 bp in length
34621: gap of unknown length
34621: contig of 864 bp in length
35157: gap of unknown length
35157: contig of 536 bp in length
36196: gap of unknown length
36196: contig of 1039 bp in length
37333: gap of unknown length
37333: contig of 1137 bp in length
38054: gap of unknown length
38054: contig of 721 bp in length
38762: gap of unknown length
38762: contig of 708 bp in length
39524: gap of unknown length
39524: contig of 762 bp in length
40176: gap of unknown length
40176: contig of 652 bp in length
42851: gap of unknown length
42851: contig of 2675 bp in length
43760: gap of unknown length
43760: contig of 909 bp in length
44291: gap of unknown length
44291: contig of 531 bp in length
44840: gap of unknown length
44840: contig of 549 bp in length
45627: gap of unknown length
45627: contig of 787 bp in length
46298: gap of unknown length
46298: contig of 671 bp in length
47167: gap of unknown length
47167: contig of 869 bp in length
47878: gap of unknown length
47878: contig of 711 bp in length
48711: gap of unknown length
48711: contig of 833 bp in length
48959: gap of unknown length
48959: contig of 248 bp in length
gap of unknown length
```



```

/chrnosome="1"
/map="p34.3-36.11"
/clone="RP1-144C9"
/clone_lib="RFCI-1"
complement(1710..2227)
/notes="match: GSS: Em:AQ355665"
complement(1780..2227)
/notes="match: GSS: Em:AQ355618"
2858..3151
/notes="match: GSS: Em:AQ627537.1"
complement(11439..11580)
/notes="match: GSS: Em:AQ031787"
complement(24560..25030)
/notes="match: GSS: Em:AQ307331"
complement(24637..25031)
/notes="match: GSS: Em:AQ234890"
27999..28360
/notes="match: STS: Em:G14651"
29391..29849
/notes="match: STS: Em:G06203"
complement(37609..37922)
/notes="match: GSS: Em:AQ114724"
47849..48201
/notes="match: GSS: Em:AQ553574.1"
63243..63496
/notes="match: GSS: Em:AQ142826"

BASE COUNT 21006 a 18901 c 18290 g 19125 t
ORIGIN

Query Match 14.4%; Score 34.4; DB 10; Length 77322;
Best Local Similarity 63.1%; Pred. No. 1.8;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 18 TGCTCTCCATCCAGGAGCCAGTGCACCTATGGGTGGGCTGGCCCTGTCCTC 77
Dn 31300 TTGTGCTTCCTCCCTCCAGGATCCCTTTGGTGAGTGTGTTCAGGATGCACCA 31241
QY 78 CTCCTGACCCCTCTGGCAGCTCA 101
Dn 31240 CTCAGATACCTTCAGGCAACACA 31217

RESULT 6
AC006445 200000 bp DNA HTG 15-SEP-1999
LOCUS Homo sapiens chromosome 4, *** SEQUENCING IN PROGRESS ***, 7
DEFINITION unordered pieces.
ACCESSION AC006445
VERSION AC006445.10 GI:5881529
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
1 (bases 1 to 200000)
Direct Submission
Unpublished
2 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
Submitted (28-JAN-1999) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
On Sep 14, 1999 this sequence version replaced gi:5822658.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

misc_feature
/chrnosome="1"
/clone="RP1-144C9"
/clone_lib="RFCI-1"
complement(1710..2227)
/notes="match: GSS: Em:AQ355665"
complement(1780..2227)
/notes="match: GSS: Em:AQ355618"
2858..3151
/notes="match: GSS: Em:AQ627537.1"
complement(11439..11580)
/notes="match: GSS: Em:AQ031787"
complement(24560..25030)
/notes="match: GSS: Em:AQ307331"
complement(24637..25031)
/notes="match: GSS: Em:AQ234890"
27999..28360
/notes="match: STS: Em:G14651"
29391..29849
/notes="match: STS: Em:G06203"
complement(37609..37922)
/notes="match: GSS: Em:AQ114724"
47849..48201
/notes="match: GSS: Em:AQ553574.1"
63243..63496
/notes="match: GSS: Em:AQ142826"

BASE COUNT 21006 a 18901 c 18290 g 19125 t
ORIGIN

Query Match 14.4%; Score 34.4; DB 41; Length 200000;
Best Local Similarity 57.7%; Pred. No. 2.2;
Matches 60; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 112 GCGCGGTATGACTTTGCACTGAAGCTGAAGGAGCTTTTCTGACAAATTCCTCTATG 171
Dn 36776 GACAGGATATCACTCTGTTGCCAAGCTAAAGTCAGTTACATGATAATGACTACTGTA 36835
QY 172 AGTCACGCTCTCTGGAATTCCTGAAATTCGCTCTCTCTCC 215
Dn 36836 ACCTCAATTCCTGGGATCAAGTATGCTCTCTCTCTCTCC 36879

RESULT 7
HSJ858B16 36676 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 858B16 on chromosome 22. Contains the
KIAA0542 gene and a gene for a novel protein similar to hamster
PSSC (Phosphatidylserine Decarboxylase Proenzyme, EC 4.1.1.65).
Contains ESTs, GSSs and a putative CpG island, complete sequence.
AL096768
ACCESSION AL096768.7 GI:5596770
VERSION HTG; CpG island; KIAA0542; Phosphatidylserine Decarboxylase
KEYWORDS Proenzyme; PSSC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 36676)
Barlow,K.
Direct Submission
Submitted (25-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 27, 1999 this sequence version replaced gi:5579004.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Eni., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C.elegans/wormpep
This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
```

feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

858B16 is from the library RPI-5 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://dacpac.med.buffalo.edu/VSCOR:pcypac2>

IMPORTANT: This sequence is not the entire insert of clone 858B16. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 694B4 (AL031255) is at 36572 in this sequence.

FEATURES
source

Location/Qualifiers

```

1..36676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/clone_lib="RPI-5"
/clone="RPI-858B16"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12171..12948,14414..14546,
14669..14747,15784..15895,15983..16218)
/genes="dJ858B16.1"
/note="match: ESTs: Em:R53362 Em:278408 Em:AA521233
Em:R33984 Em:AI808913 Em:AA158770 Em:AA836285 Em:AA974403
Em:AA594535 Em:AI858196 Em:AI425026 Em:R51102 Em:AA582394
Em:AA829254 Em:AA553345 Em:AI809747 Em:AA846853
Em:AA971234 Em:R54651 Em:AI141022 Em:AA159572 Em:AI361548
Em:AA916775 Em:AI191287 Em:AV031353 Em:AA779744 Em:N64034
Em:AV075750 Em:AV095355 Em:AA865102 Em:AI232370
Em:AA595561 Em:AA088219 Em:W25871 Em:N56896 Em:W90666
Em:AV030271 Em:AV116402 Em:AV122327 Em:AI024249
Em:AA673371 Em:AI130893 Em:AI019759 Em:AI770152
Em:AA149582 Em:N30173 Em:AI342072 Em:R54847 Em:AI126154
Em:AA181858 Em:AA044074 Em:N54928 Em:AA215582 Em:AA398600
Em:AI623440 Em:AI024124 Em:AA637284 Em:AI824779
Em:AI094871 Em:AI192647 Em:N63934 Em:AV083921"
/evidence="not_experimental"
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..12948,14414..14546,14669..14747,
15784..15895,15983..16218)
/genes="dJ858B16.1"
/note="match: cDNAs: Em:AB011114"
/evidence="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
273..16218
/genes="dJ858B16.1"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11524,12171..12948,14414..14546,
14669..14747,15784..15895,15983..16087)
/genes="dJ858B16.1"
/note="different from published protein but supported by
GENES and GENSCAN"
/codon_start=1
/evidence="not_experimental"
/product="dJ858B16.1.1 (KIAA0542 (isoform 1))"
/protein_id="CAB56392.1"
/db_xref="GI:5921489"
/translation="QAILHAERQLLYRSFWMHQAAARHQBQEWQTVACAHHRHGRGL
KFAFLMRSAGQLRTYGRVRAAEFHMAQLLRWANSQWRECLALRGAEKRLMRAD
LHHQSVLHRLAQAWTYOGRVRSILREVAARESHNNRQLLRGALRWKENTMARVDE
AKKTFQASTHYRTICSKVLQVREAVSVOMTYRQEDCAIWEAQKVLDRGCLRTWFO
RWDCSSRAQRLQLERAVOHHRQLLLEGLARKWTHLQCVKRLLRHQSTQLLAQ
RLSRTCFRWRQQLAARQEQATVRALWFAFSLQAKVWATLWAFVLERRKKARLQ
WALQAYQGLLQEGATRLLRFAASRKASRQOLQAOQVQAHSLSHRAVRRCATLKKOK
VLGRGKRPOLAAIAPSRKVTFFGPLNRIAGAAGDTLETKRQASRPLGALGLAA
EEPHALEL"

```

mRNA

gene

CDS

```

PFATGPELLLLPLSFMPGCAAPARVSAQRATPRDKPPVPSLSASVDPDPLLPGD
FSATRAPGSLTAGSLDELELEIQQLLHYTTKQNLNSCRQASRRWLELNRE
EPGEDVEQVQKLEQVEMQIQLABELAQRPICACVARIQALCALC"
join(<273..454,1413..1488,1978..2073,2538..2614,
3997..4099,5606..5702,8812..8972,9435..9509,10811..10885,
11023..11199,11271..11528)
/genes="dJ858B16.1"
/note="Original published form; could be artefactual as it
is result of a large intron read-through; match: proteins:
Tr:060289"
/codon_start=1
/evidence="not_experimental"
/product="dJ858B16.1.2 (KIAA0542 (isoform 2))"
/protein_id="CAB56393.1"
/db_xref="GI:5921490"
/translation="QAILHAERQLLYRSFWMHQAAARHQBQEWQTVACAHHRHGRGL
KFAFLMRSAGQLRTYGRVRAAEFHMAQLLRWANSQWRECLALRGAEKRLMRAD
LHHQSVLHRLAQAWTYOGRVRSILREVAARESHNNRQLLRGALRWKENTMARVDE
AKKTFQASTHYRTICSKVLQVREAVSVOMTYRQEDCAIWEAQKVLDRGCLRTWFO
RWDCSSRAQRLQLERAVOHHRQLLLEGLARKWTHLQCVKRLLRHQSTQLLAQ
RLSRTCFRWRQQLAARQEQATVRALWFAFSLQAKVWATLWAFVLERRKKARLQ
WALQAYQGLLQEGATRLLRFAASRKASRQOLQAOQVQAHSLSHRAVRRCATLKKOK
VLGRGKRPOLAAIAPSRKVTFFGPLNRIAGAAGDTLETKRQASRPLGALGLAA
EEPHALEL"
620..913
/note="AluX repeat: matches 1..294 of consensus"
1444..1687
/genes="dJ858B16.1"
/note="match: GSS: Em:AAQ54234"
3250..3536
/note="AluJo repeat: matches 2..302 of consensus"
3634..3836
/note="L2 repeat: matches 2497..2701 of consensus"
4356..4646
/note="AluX repeat: matches 21..312 of consensus"
4650..4754
/note="AluSg/x repeat: matches 200..307 of consensus"
4990..5281
/note="AluX repeat: matches 1..292 of consensus"
5806..6231
/note="MLTIC repeat: matches 6..460 of consensus"
6529..6668
/note="FRAM repeat: matches 1..140 of consensus"
6671..6966
/note="AluJ repeat: matches 1..295 of consensus"
6985..7081
/note="LTR40a repeat: matches 1..100 of consensus"
7929..8097
/note="MIR repeat: matches 82..255 of consensus"
9766..10068
/note="AluJ repeat: matches 1..301 of consensus"
10089..10096
/note="14 copies 2 mer ta 100% conserved"
10104..10399
/note="AluJB repeat: matches 1..293 of consensus"
12099..12259
/note="MIR repeat: matches 46..208 of consensus"
13063..13779
/note="L2 repeat: matches 1570..2750 of consensus"
15108..15404
/note="AluSg repeat: matches 1..295 of consensus"
complement(16160)
/genes="dJ858B16.2"
complement(join(16160..17505,18222..18382,18665..18811,
19003..19141,19318..19554,21353..21518,23432..23685,
28473..28494))
/genes="dJ858B16.2"
/note="match: cDNAs: Em:M62722 Em:AL050371 Em:AF086277
Em:AF035304 Em:U82163 Em:AB011114; match: ESTs:
Em:AA324939 Em:AA682112 Em:AA166445 Em:AA166440
Em:AA260704 Em:AI019759 Em:AL039237 Em:AA118158
Em:AV083921 Em:AI232370 Em:AU035550 Em:AV122327
Em:AA717291 Em:AA326672 Em:AU051210 Em:AA298268

```

Em:AA144920 Em:AA673371 Em:R0603 Em:AA297077 Em:AA237290
 Em:AA615772 Em:AA297754 Em:AA297076 Em:AA636369
 Em:AI094871 Em:AA298523 Em:AA149582 Em:AA836285
 Em:AV031353 Em:AA297159 Em:AI191287 Em:AA916775 Em:R51102
 Em:N30173 Em:AA521233 Em:AA044074 Em:AI361548 Em:H69648
 Em:AA971234 Em:W00857 Em:AI623440 Em:AA553345 Em:H93993
 Em:AA046738 Em:AA637284 Em:AA151798 Em:W69452 Em:AA297086
 Em:W69536 Em:TI0522 Em:AI024124 Em:H71693 Em:AA302951
 Em:N98280 Em:N63934 Em:AA215768 Em:R94341 Em:AA177542
 Em:R48871 Em:AI342072 Em:W90666 Em:AA255719 Em:AI024249
 Em:AA829254 Em:W69460 Em:AA595561 Em:W69544 Em:T77685
 Em:AA779744 Em:N64034 Em:AV075750 Em:H38339 Em:AV095355
 Em:AV116402 Em:AI425026 Em:AA974403 Em:AA908164
 Em:AA088219 Em:AA298551 Em:AA56896 Em:AV030271 Em:AI428571
 Em:R80913 Em:AI014094 Em:H53175 Em:AI809747 Em:H38743
 Em:H62728 Em:W03262 Em:R52264 Em:AA044185 Em:R82239
 Em:AA369042 Em:AA393183 Em:AI126154 Em:R54847 Em:AA158770
 Em:AA825239 Em:R53399 Em:AI770152 Em:AA594535 Em:AA582394

Query Match 14.2%; Score 34; DB 11; Length 36676;
 Best Local Similarity 54.9%; Pred. No. 2.3;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 Qy 19 GTCCTCCATCCAGCGGAGTGGCCACTATGGGTCTGGGTGCCCCCTTGTCTCC 78
 Db 14428 GGCCTACTCTAGGATAGAGCCCCCGTCCCTCATCCCTGGCCAGTGCCTGACCCCA 14487
 Qy 79 TCTTGACCTCCTTGGCAGCTCACATGGAACAGCGCGGTATGACTTTTGCACACTGAAGC 138
 Db 14488 TCTACTCCTCTGGGAGCTTCTCAGCCACCAGGCGTGGCGCTGCACTTCAACTGCAGG 14547
 Qy 139 TG 140
 Db 14548 TG 14549

RESULT 8
 AC016298 62181 bp DNA HTG 24-NOV-1999
 LOCUS Homo sapiens clone RP11-655M17, LOW-PASS SEQUENCE SAMPLING.
 DEFINITION AC016298
 ACCESSION AC016298
 VERSION 1 GI:6467036
 KEYWORDS HTG; HTGS_PHASE0.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 62181)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Bouckgaier, B.,
 Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
 Cooke, P., DeArrelano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
 Ferrelira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gardyna, S., Grant, G., Hacos, B., Heaford, A., Horton, L.,
 Howland, J.C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lechoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Tesfaye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wyman, D., Ye, W.J., Zimmer, A. and Zody, M.
 Direct Submission
 TITLE Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
 JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RN/RepeatMasker.html
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submission@genome.wi.mit.edu
 ----- Project Information
 Center Project name: L4315
 Center clone name: 655_M_17

* NOTE: This record contains 78 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 775: contig of 775 bp in length
 * 776 1583: contig of 808 bp in length
 * 1584 2371: contig of 788 bp in length
 * 2372 3185: contig of 814 bp in length
 * 3186 3980: contig of 795 bp in length
 * 3981 4778: contig of 798 bp in length
 * 4779 5584: contig of 806 bp in length
 * 5585 6392: contig of 808 bp in length
 * 6393 7189: contig of 787 bp in length
 * 7190 7994: contig of 805 bp in length
 * 7995 8802: contig of 808 bp in length
 * 8803 9603: contig of 801 bp in length
 * 9604 10400: contig of 797 bp in length
 * 10401 11206: contig of 806 bp in length
 * 11207 12024: contig of 818 bp in length
 * 12025 12840: contig of 816 bp in length
 * 12841 13637: contig of 797 bp in length
 * 13638 14440: contig of 803 bp in length
 * 14441 15234: contig of 794 bp in length
 * 15235 16031: contig of 797 bp in length
 * 16032 16836: contig of 805 bp in length
 * 16837 17615: contig of 779 bp in length
 * 17616 18421: contig of 806 bp in length
 * 18422 19233: contig of 812 bp in length
 * 19234 20042: contig of 809 bp in length
 * 20043 20856: contig of 814 bp in length
 * 20857 21657: contig of 801 bp in length
 * gap of unknown length

*	21658	22462:	contig of 805 bp in length	
*			gap of unknown length	
*	22463	23262:	contig of 800 bp in length	
*			gap of unknown length	
*	23263	24065:	contig of 803 bp in length	
*			gap of unknown length	
*	24066	24866:	contig of 801 bp in length	
*			gap of unknown length	
*	24867	25673:	contig of 807 bp in length	
*			gap of unknown length	
*	25674	26475:	contig of 802 bp in length	
*			gap of unknown length	
*	26476	27270:	contig of 795 bp in length	
*			gap of unknown length	
*	27271	28075:	contig of 805 bp in length	
*			gap of unknown length	
*	28076	28817:	contig of 742 bp in length	
*			gap of unknown length	
*	28818	29613:	contig of 796 bp in length	
*			gap of unknown length	
*	29614	30416:	contig of 803 bp in length	
*			gap of unknown length	
*	30417	31207:	contig of 791 bp in length	
*			gap of unknown length	
*	31208	32021:	contig of 814 bp in length	
*			gap of unknown length	
*	32022	32825:	contig of 804 bp in length	
*			gap of unknown length	
*	32826	33591:	contig of 766 bp in length	
*			gap of unknown length	
*	33592	34403:	contig of 812 bp in length	
*			gap of unknown length	
*	34404	35213:	contig of 810 bp in length	
*			gap of unknown length	
*	35214	36022:	contig of 809 bp in length	
*			gap of unknown length	
*	36023	36829:	contig of 807 bp in length	
*			gap of unknown length	
*	36830	37623:	contig of 794 bp in length	
*			gap of unknown length	
*	37624	38411:	contig of 788 bp in length	
*			gap of unknown length	
*	38412	39213:	contig of 802 bp in length	
*			gap of unknown length	
*	39214	40010:	contig of 797 bp in length	
*			gap of unknown length	
*	40011	40741:	contig of 731 bp in length	
*			gap of unknown length	
*	40742	41570:	contig of 829 bp in length	
*			gap of unknown length	
*	41571	42298:	contig of 728 bp in length	
*			gap of unknown length	
*	42299	42980:	contig of 682 bp in length	
*			gap of unknown length	
*	42981	43786:	contig of 806 bp in length	
*			gap of unknown length	
*	43787	44584:	contig of 798 bp in length	
*			gap of unknown length	
*	44585	45392:	contig of 808 bp in length	
*			gap of unknown length	
*	45393	46206:	contig of 814 bp in length	
*			gap of unknown length	
*	46207	46991:	contig of 785 bp in length	
*			gap of unknown length	
*	46992	47813:	contig of 822 bp in length	
*			gap of unknown length	
*	47814	48623:	contig of 810 bp in length	
*			gap of unknown length	
*	48624	49437:	contig of 814 bp in length	
*			gap of unknown length	
*	49438	50183:	contig of 746 bp in length	
*			gap of unknown length	
*	50184	50985:	contig of 802 bp in length	

* * *	50986	51800:	gap of unknown length	14.1%;	Score 33.8;	DB 44;	Length 62181;
* * *			contig of 815 bp in length	Best Local Similarity	53.4%;	Pred. No. 2.7;	
* * *	51801	52595:	gap of unknown length	Matches 71;	Conservative	0;	Mismatches 62; Indels 0; Gaps 0;
* * *			contig of 795 bp in length				
* * *	52596	53395:	gap of unknown length				
* * *			contig of 800 bp in length				
* * *	53396	54205:	gap of unknown length				
* * *			contig of 810 bp in length				
* * *	54206	54996:	gap of unknown length				
* * *			contig of 791 bp in length				
* * *	54997	55783:	gap of unknown length				
* * *			contig of 787 bp in length				
* * *	55784	56587:	gap of unknown length				
* * *			contig of 804 bp in length				
* * *	56588	57389:	gap of unknown length				
* * *			contig of 802 bp in length				
* * *	57390	58203:	gap of unknown length				
* * *			contig of 814 bp in length				
			gap of unknown length				

Query Match	
Best Local Similarity	53.4%;
Matches 71;	Conservative

Qy	10	GACTTCAGTGTCTCTCATCCGACGAGCGGACGAGTGGCCACTATGGGTCTGGGGTCCCC	69
Db	11905	GATTTTAAATCTCTTCTACCCCTGAAGCTGGGAGTGGTCTCTCTTTGGTTGGGCTCCCC	11964
Qy	70	TTGTCTCTCTTTGACCTCTCTTTGGCAGCTCACAATGGAACAGGCGCCGGTATGACTTTGC	129
Db	11965	TCCTTTCTTTTGGCCCTGGGTTGGCAGTCCCTCGAAAAGGCTCTACATCCCTTGC	12024
Qy	130	AACTGAAGCTGAA	142
Db	12025	GCCTGGGTTGCA	12037

RESULT	9		
HSJ773A18/c			
LOCUS			
DEFINITION			
HSJ773A18 128379 bp DNA	PRI	23-NOV-1999	
Human DNA sequence from clone 773A18 on chromosome lp13.2-21.1			
Contains genes for RAS-RELATED PROTEIN RAP-1A (C21KG) (KREV-1			
PROTEIN) (GTP-BINDING PROTEIN SMG-P21A) (G-22K), KCND3 (potassium			
voltage-gated channel, Shal-related subfamily, member 3), PROBABLE			
ATP-DEPENDENT RNA HELICASE P47 HOMOLOG, ESTs, STSS, GSSs and CpG			
Islands complete sequence.			
AL049357			
AL049357.19	GI:5596765		
HTG: C21KG; CpG island; G-22K; GTP-BINDING PROTEIN; KCND3; KREV-1			
PROTEIN; RAP1A; RAS-RELATED PROTEIN RAP1-A; SMG-P21A.			
human.			
SOURCE			
ORGANISM			
Human sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
Eutheria; Primates; Catarrhini; Hominidae; Homo.			
1 (bases 1 to 128379)			
Corvillie G.			
Direct Submission			
Submitted (03-SEP-1999) Sanger Centre, Hinxton, Cambridgeshire,			
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone			
requests: clonerequest@sanger.ac.uk			
On Jul 27, 1999 this sequence version replaced gi:5579000.			
During sequence assembly data is compared from overlapping clones.			
where differences are found these are annotated as variations			
together with a note of the overlapping clone name. Note that the			
variation annotation may not be found in the sequence submission			
corresponding to the overlapping clone, as we submit sequences with			
only a small overlap as described above.			
The following abbreviations are used to associate primary accession			
numbers given in the feature table with their source databases:			
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information			
on the WORMPEP database can be found at			
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence is			
the entire insert of clone 773A18. This sequence has been finished			


```

repeat_region /note="L1MEC repeat: matches 1219. .1304 of consensus"
20187. .20493
repeat_region /note="AluJb repeat: matches 1. .311 of consensus"
20494. .20653
repeat_region /note="L1MEC repeat: matches 1304. .1454 of consensus"
20747. .21111
repeat_region /note="MLT1J repeat: matches 111. .516 of consensus"
21166. .21597
repeat_region /note="L1MEC repeat: matches 1509. .1991 of consensus"
21633. .21941
repeat_region /note="AluJb repeat: matches 3. .308 of consensus"
21945. .22194
repeat_region /note="L1M4 repeat: matches 3044. .3318 of consensus"
22225. .22726
misc_feature complement(22225. .22726)
repeat_region /note="match: GSS: Em:AQ180467"
22604. .22705
repeat_region /note="FRAM repeat: matches -2. .120 of consensus"
22908. .23214
repeat_region /note="AluXx repeat: matches 1. .310 of consensus"
23215. .23635

Query Match 14.1%; Score 33.6; DB 11; Length 128379;
Best Local Similarity 48.9%; Pred. No. 3.3;
Matches 90; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 8 GGGACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGCTGCC 67
Db 97429 GTGACCTCTTTTCTCCACCCCTGGTCTCAGGAGGAGCAGGTGGCTTTCTTC 97370

QY 68 CTTGTCTCCTCTTGACCTCTTGCCAGCTCACATGGAACAGGCGCGGTATGACTTT 127
Db 97369 CTTCTCTAATATGCTCTCCCTGGCAAGCTGCCAACCACTGCTCTGTACAGCTA 97310

QY 128 GCAACTCAAGCTGAAGGAGCTTTTCTGACAAATCTCTCTATGATCCAGCTTCCGGA 187
Db 97309 CTAAGTGAGATGGCCAGTGTATGTATAAATGACCTTTCTCTGCTGATCCAGCC 97250

QY 188 ATTG 191
Db 97249 ATGG 97246

RESULT 10
AC014817/c
LOCUS AC014817 26065 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC014817
VERSION AC014817.1 GI:6436518
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 26065)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210440 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1. .26065
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 7321 a 5960 c 5929 g 6855 t
ORIGIN

```

```

Query Match 14.1%; Score 33.6; DB 43; Length 26065;
Best Local Similarity 59.4%; Pred. No. 3;
Matches 57; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 5 ACCGGAGCTTCAAGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGTCTGGCT 64
Db 1415 ACAGGCCATGATGCTTGTCTCCAGCGCTCATGACGCTTTTGGCTCGCT 1356

QY 65 GCGCTTGCTCTCTCTTGACCTCTCTGGCAGCTC 100
Db 1355 GCGCTAAATCCGCCACCCCTCTCTGGCAGCTC 1320

RESULT 11
AC017080/c
LOCUS AC017080 106791 bp DNA HTG 09-DEC-1999
DEFINITION Homo sapiens clone RP11-469G4, *** SEQUENCING IN PROGRESS ***, 34
unordered pieces.
ACCESSION AC017080
VERSION AC017080.1 GI:6554067
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 106791)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 106791)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT Center project name: H.NH0469G04.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2077: contig of 2077 bp in length
2078 2097: gap of unknown length
6097 6117: contig of 4000 bp in length
6118 8806: contig of 2689 bp in length
8807 8826: gap of unknown length
8827 11355: contig of 2729 bp in length
11356 11576: gap of unknown length
11576 13852: contig of 2277 bp in length
13853 13872: gap of unknown length
13873 16840: contig of 2968 bp in length
16841 16860: gap of unknown length
16861 19354: contig of 2494 bp in length
19355 19373: gap of unknown length
19374 22201: contig of 2827 bp in length
22201 26219: contig of 4000 bp in length
26220 26238: gap of unknown length
26239 28539: contig of 2301 bp in length
28540 28558: gap of unknown length
28559 31633: contig of 3075 bp in length
31634 35652: gap of unknown length
35653 35671: gap of unknown length
35672 39671: contig of 4000 bp in length
39691 42315: contig of 2625 bp in length
42316 42334: gap of unknown length

```



```
* 116978 127321: contig of 10354 bp in length
* 127332 127349: gap of unknown length
* 127350 138238: contig of 10889 bp in length
* 138239 138256: gap of unknown length
* 138257 151790: contig of 13534 bp in length
* 151791 151808: gap of unknown length
* 151809 164982: contig of 13174 bp in length
* 164983 165000: gap of unknown length
* 165001 184895: contig of 19895 bp in length
* 184896 184913: gap of unknown length
* 184914 211190: contig of 26277 bp in length.
      Location/Qualifiers
        1..211190
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /clone="RP11-549G13"
BASE COUNT 52817 a 52007 c 53208 g 52618 t 540 others
ORIGIN

Query Match      14.0%; Score 33.4; DB 44; Length 211190;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 62 GCTGCCCTTGTCTCTCTGACCTCTTGGAGCTCAGTGGACAGGCGCGGTAT 121
   || || || || || || || || || || || || || || || || || || ||
Db 162017 GCAGAACCTTGTCTCTCTCCCGACCCACTGGAGCTGGCAGACAAAGCCCGAGTGGAG 161958

QY 122 GACTTTTGCACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCTATGATCCAGCTT 181
   || || || || || || || || || || || || || || || || || || ||
Db 161957 CCAGGCTTCTCAGAGGGCATGTGACTCACCTGTGCTCTCCACCCCTGCCAGCTG 161898

QY 182 CCTGGAATGCTGAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
   || || || || || || || || || || || || || || || || || || ||
Db 161897 CAACCCCTGCAAGAAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 161854

RESULT 13
AC016764
LOCUS      Homo sapiens clone Rpl11-549G13, WORKING DRAFT SEQUENCE, 1 unordered
DEFINITION      Homo sapiens clone Rpl11-549G13, WORKING DRAFT SEQUENCE, 1 unordered
pieces
AC016764
VERSION      AC016764.2 GI:6554041
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 191010)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 191010)
Waterston,R.H.
Direct Submission
TITLE      Submitted (04-DEC-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT      On Dec 10, 1999 this sequence version replaced gi:6524275.
Center project name: H.NH0549G13.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 191010: contig of 191010 bp in length.
      Location/Qualifiers
        1..191010
          /organism="Homo sapiens"
```

```
/db_xref="taxon:9606"
/clone="RP11-549G13"
BASE COUNT 46669 a 48592 c 47516 g 47765 t 468 others
ORIGIN

Query Match      14.0%; Score 33.4; DB 45; Length 191010;
Best Local Similarity 50.0%; Pred. No. 3.9;
Matches 82; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 62 GCTGCCCTTGTCTCTCTGACCTCTTGGAGCTCAGTGGACAGGCGCGGTAT 121
   || || || || || || || || || || || || || || || || || || ||
Db 148271 GCAGAACCTTGTCTCTCTCCCGACCCACTGGAGCTGGCAGACAAAGCCCGAGTGGAG 148330

QY 122 GACTTTTGCACTGAAGCTGAAGGAGTCTTTTCTGACAAATTCCTCTATGATCCAGCTT 181
   || || || || || || || || || || || || || || || || || || ||
Db 148331 CGAGGCTTCTCAGAGGGCATGTGACTCACCTGTGCTCTCTCCACCCCTGCCAGCTG 148390

QY 182 CCTGGAATGCTTGAAGTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
   || || || || || || || || || || || || || || || || || || ||
Db 148391 CAACCCCTGCAAGAAAGTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 148434

RESULT 14
OSU25430
LOCUS      Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete
DEFINITION      Oryza sativa nucleotide pyrophosphatase precursor, mRNA, complete
cds.
ACCESSION      U25430
VERSION      U25430.1 GI:818848
KEYWORDS      rice.
SOURCE      Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (sites)
AUTHORS      Hsing,Y.C., Tsao,C.V., Chow,T., Hsieh,J. and Chen,Z.
TITLE      Rice early embryogenesis gene
JOURNAL      Unpublished
REFERENCE 2 (bases 1 to 1801)
AUTHORS      Hsing,Y.C.
TITLE      Direct Submission
JOURNAL      Submitted (20-APR-1995) Yue-le C. Hsing, Academia Sinica, Institute
of Botany, Yien-chu-yuan Rd. Taipei, Taiwan, 11529, Republic of
China
FEATURES
    Location/Qualifiers
        1..1801
          /organism="Oryza sativa"
          /strain="Tainung 67"
          /db_xref="taxon:4530"
          /clone="OSE4"
          /dev_stage="3 to 5 days after anthesis"
          /clone_lib="constructed by Y.C. Hsing et al., Academia
          Sinica, Institute of Botany, Taipei, Taiwan, 11529,
          Republic of China"
          /note="early embryos in seeds"
          /note="signal peptide to endoplasmic reticulum"
          /EC_number="3.6.1.9"
          /note="nucleotide phosphodiesterase"
          /codon_start=1
          /product="nucleotide pyrophosphatase precursor"
          /protein_id="AAA67067.1"
          /db_xref="GI:818849"
          /translation="MAAATAAAPPFAAGDSPPTALLPRTTTTAGAPAPRPSAS
          SRLHLTAALAVATSTLLLPRTPLSAAPAPAPAPAAQAQKLEKPVVILLISDGRF
          GYHQKATPHIIRLIGNGTSAATGLVPIPTLTFFNNYSIATGLYPSHGINNFDP
          PISDYFTMSHSEPKWLGEPPLWATAADGTOAATYFPGSGVKSGNSDCPDYCRHY
          NGSVPFERVDAIILGYFDLPDSEMPQFLTFYEDPDHQHGVGPDPAITVAVRIDE
          MIGRLTAGLERGVFDVNVLLVGDHGMVGTCDKRLVLDLAPWIKLEEDWVLSMTP
          LLAIKRPDDMSLPDVAKMGNEGIGSGKVENGEYLRWYKEDLPKRLHYADSVRIPII
```

GLPEGYKVKRSDKNECGAHGYNAPFMSRTIFIAHGRFEGGRVPSFENVIELY
NVIASILNLEPAPNGSSSFPDILLPSE"

mat_peptide 97. .1476 /EC_number="3.6.1.9"
/note="nucleotide phosphodiesterase"
polyA_site 1484. .1488 /product="nucleotide pyrophosphatase"
polyA_site 1572. .1576 /note="putative"
polyA_site 1572. .1576 /note="putative"
BASE COUNT 374 a 527 c 447 g 453 t
ORIGIN

Query Match 13.8%; Score 33; DB 7; Length 1801;
Best Local Similarity 47.5%; Pred. No. 4.1;
Matches 96; Conservative 0; Mismatches 106; Indels 0; Gaps 0;
QY 21 CTCCTCCATCCGAGGCGCAGTGGCCACTATGGGCTCGGGTGCCTCTTCTCCTC 80
Db 181 CACCTCTCTCTACCGCGCGTGGCGTGCCTCTCTCTCTCTCTCTCTCTCTCT 240
QY 81 TTGACCTCTCTTGGCAGTTCACATGAACAGCGCGGTATGACTTTGCAACTGAAGCTG 140
Db 241 CGACGCGCTCTCCGCGCGCGCGCTCTGGCGCGCGCGCGCGCGAGTGAAGCTC 300
QY 141 RAGGAGTCTTTCTGACAAATCTCTCTATGAGTCCAGCTTCTCGGATGCTTGAANA 200
Db 301 GAGAGCGGTGTGTGATCTCTATCTCTCGAGCGGTCTCCGCTCGGGTACCAGACAAG 360
QY 201 TCTGCT 222
Db 361 GCGCGACCGCGCACATCCACC 382

RESULT 15
HS591N18 97580 bp DNA PRI 23-NOV-1999
LOCUS Human DNA sequence from clone 591N18 on chromosome 22q13.1-13.2.
DEFINITION Contains a COX6B (Cytochrome C Oxidase subunit VIB (EC 1.9.3.1))
pseudogene, ESTs, GSSs and two putative CpG islands, complete
sequence.
ACCESSION AL031594
VERSION AL031594.9 GI:5050980
KEYWORDS HWG; COX6B; CpG Island; Cytochrome C Oxidase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 97580)
AUTHORS Steward,C.
TITLE Direct Submission
JOURNAL Submitted (25-JUN-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humqu@rsb.sanger.ac.uk
requests: clonerequest@sanger.ac.uk
COMMENT On Jun 11, 1999 this sequence version replaced gi:4914529.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TREMBL
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 22, constructed by the Sanger Centre Chromosome 22

Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr22
591N18 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: PCYPAC2
IMPORTANT: This sequence is not the entire insert of clone 591N18.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 229A8 (286090) is at 97475 in this
sequence. The true right end of clone 1042K10 (AL022238) is at
34197 in this sequence.

FEATURES
source 1. 97580
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="q13.1-13.2"
/clone="RP4-591N18"
/clone_lib="RPC1-4"
1. 227
repeat_region /note="AluY repeat: matches 75. .300 of consensus"
996. .1104
repeat_region /note="FLAM_C repeat: matches 1. .108 of consensus"
1147. .1220
repeat_region /note="AluJ/FRAM repeat: matches 217. .294 of consensus"
1225. .1501
repeat_region /note="AluJo repeat: matches 12. .282 of consensus"
1552. .1864
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
1869. .2134
repeat_region /note="AluJb repeat: matches 21. .275 of consensus"
2942. .3234
repeat_region /note="AluSc repeat: matches 4. .295 of consensus"
4854. .5144
repeat_region /note="AluJb repeat: matches 1. .308 of consensus"
6310. .6584
misc_feature /note="match: GSS AQ038173"
6506. .6575
repeat_region /note="MER3 repeat: matches 134. .209 of consensus"
6585. .6665
repeat_region /note="FLAM_A repeat: matches 27. .116 of consensus"
6673. .6894
repeat_region /note="MER3 repeat: matches 148. .167 of consensus"
6695. .6988
repeat_region /note="AluSg repeat: matches 1. .290 of consensus"
6989. .7068
repeat_region /note="MER3 repeat: matches 58. .148 of consensus"
7497. .7761
repeat_region /note="AluJb repeat: matches 1. .301 of consensus"
7845. .8136
repeat_region /note="AluSx repeat: matches 1. .299 of consensus"
8467. .8776
repeat_region /note="AluY repeat: matches 1. .311 of consensus"
9209. .9505
repeat_region /note="AluSg repeat: matches 1. .295 of consensus"
11780. .12007
repeat_region /note="L1ME1 repeat: matches 5465. .5691 of consensus"
12018. .12387
repeat_region /note="HWRV23 repeat: matches 1184. .1573 of consensus"
12454. .12544
repeat_region /note="L1ME1 repeat: matches 5679. .5763 of consensus"
12545. .12843
repeat_region /note="AluJb repeat: matches 1. .297 of consensus"
12844. .13179
repeat_region /note="L1ME1 repeat: matches 5763. .6159 of consensus"
13240. .13373
repeat_region /note="AluJo repeat: matches 1. .136 of consensus"
13408. .13612
repeat_region /note="AluSg/x repeat: matches 96. .293 of consensus"
13625. .13924
repeat_region /note="AluSx repeat: matches 1. .301 of consensus"
13936. .14230
repeat_region

```
repeat_region /note="AluSg repeat: matches 1. .297 of consensus"
14237. .14466
repeat_region /note="MIR repeat: matches 11. .261 of consensus"
14729. .14785
repeat_region /note="MIR repeat: matches 202. .261 of consensus"
15147. .15467
repeat_region /note="L2 repeat: matches 2370. .2690 of consensus"
15625. .15910
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
16000. .16300
repeat_region /note="AluSg repeat: matches 1. .291 of consensus"
16422. .16732
repeat_region /note="AluJb repeat: matches 1. .311 of consensus"
16733. .16813
misc_feature /note="L2 repeat: matches 2667. .2746 of consensus"
complement(16814. .17293)
repeat_region /note="match: GSSs AQ377982 AQ380087"
16907. .17100
repeat_region /note="MIR repeat: matches 2. .197 of consensus"
18063. .18376
repeat_region /note="AluJb repeat: matches 1. .307 of consensus"
18462. .18537
repeat_region /note="38 copies 2 mer tt 78% conserved"
18540. .19522
misc_feature /note="SVA repeat: matches 2. .954 of consensus"
19061. .20976
repeat_region /note="random repeat. Some base discrepancies edited.
Assembly is consistent with Restriction Digest."
19115. .20959
misc_feature /note="CpG island"
/evidence=not_experimental
19540. .20001
repeat_region /note="SVA repeat: matches 521. .954 of consensus"
20019. .20360
repeat_region /note="SVA repeat: matches 521. .845 of consensus"
20361. .20854
misc_feature /note="SVA repeat: matches 519. .994 of consensus"
complement(20855. .20945)
repeat_region /note="match: GSS AQ473058"
21493. .21658
repeat_region /note="AluSx repeat: matches 1. .165 of consensus"
21793. .22086
repeat_region /note="AluSx repeat: matches 1. .297 of consensus"
22379. .22534
repeat_region /note="FRAM repeat: matches -1. .162 of consensus"
23688. .24049
repeat_region /note="MLT1A1 repeat: matches 1. .365 of consensus"
24100. .24408
repeat_region /note="AluYb8 repeat: matches 1. .305 of consensus"
25227. .25536
repeat_region /note="AluJb repeat: matches 1. .310 of consensus"
25556. .25779
repeat_region /note="AluJb repeat: matches 58. .297 of consensus"
25834. .26147
repeat_region /note="AluSg repeat: matches 1. .307 of consensus"
26148. .26192
repeat_region /note="MIR repeat: matches 187. .230 of consensus"
26377. .26688
repeat_region /note="AluSx repeat: matches 2. .311 of consensus"
26884. .27010
repeat_region /note="FRAM_C repeat: matches 1. .133 of consensus"
30237. .30344
repeat_region /note="MIR repeat: matches 103. .206 of consensus"
31467. .31777
repeat_region /note="AluSg repeat: matches 1. .310 of consensus"
31963. .32080
repeat_region /note="AluJb repeat: matches 1. .118 of consensus"
32118. .32408
repeat_region /note="AluJb repeat: matches 1. .292 of consensus"
32623. .32927
repeat_region /note="AluJb repeat: matches 1. .300 of consensus"
33051. .33354
repeat_region /note="AluSg repeat: matches 1. .304 of consensus"
```

```
repeat_region 33381. .33519
/note="L2 repeat: matches 2572. .2709 of consensus"
repeat_region 33560. .33777
/note="AluSx repeat: matches 1. .219 of consensus"
repeat_region 34284. .34597
/note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 35062. .35370
/note="AluJb repeat: matches 1. .309 of consensus"
repeat_region 35600. .35753
/note="MIR repeat: matches 2. .230 of consensus"
repeat_region 35782. .35854
/note="MER21B repeat: matches 720. .790 of consensus"
repeat_region 35912. .36626
/note="MER21B repeat: matches 1. .713 of consensus"
repeat_region 36650. .36969
/note="AluY repeat: matches 1. .308 of consensus"
repeat_region 37152. .37273
/note="AluSg/x repeat: matches 185. .302 of consensus"

Query Match 13.8%; Score 33; DB 11; Length 97580;
Best Local Similarity 55.3%; Pred. No. 5;
Matches 63; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 91 TTGGCAGCTCACATGGAACAGGCCGGGTATGACTTGCACCTGAACCTGAAGAGTCTT 150
Db 2606 TTCCCACTCCACATCCCGCGTTTAGTAAGACCTTTTCAAAAGCAAGGTAGGTGAAC 2665
QY 151 TTCGACAAATTCCTCTATGAGTCCAGCTTCTCGAAATTCCTTGAATAATCTG 204
Db 2666 TGTAGAAGAGTTCATGTTATAAGTACCGCAACTTTGTATTCTCTGAAGTCTGTG 2719
```

Search completed: March 21, 2000, 14:39:02
Job time: 3557 sec

181 TCCTGGAATTGCTTGAAA-NTCTGCCCTCCTCCTCCATCTCCCTTCAGGGACCAGCGTCA


```
QY 33 AGGAGCGCAGTGGCCACTATGGGCTGTGGGCTGGCCCTTGTCTCTCTCTTGTGACCTCTCT 92
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 ACATCGCGGGCGCGCGCTATAGGAGCGCTGTGGCCCTTGTCTCTCTCTCTCTCTCTCTCT 771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 93 GGCAGCTCATCGGAACAGG 112
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 TTCTGCTGTACAGGAAAGG 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
T18696
ID T18696 standard; cDNA; 3510 BP.
AC T18696;
DE 05-JUL-1996 (first entry)
DE RAP-1 radiation protecting checkpoint protein cDNA.
KW RAP-1; radiation protecting checkpoint protein; apoptosis;
KW cell death; cancer; diagnosis; therapy; radiotherapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 176..2122
FT FT /*tag= a
FT misc_rna 31077..3510
FT FT /*tag= b
FT FT /label= cDNA3
FT misc_difference 3648
FT FT /*tag= b
FT FT /note= "base n at position 3648 is not identified
FT FT in the specification"
PN WO9611562-A2.
PD 25-APR-1996.
PF 11-OCT-1995; U12445.
PR 11-OCT-1994; IL-111238.
PA (SHOS/) SHOSHAN H Z.
PA (UYRA-) UNIV RAMOT APPLIED RES & IND DEV LTD.
PI Canaanani D.
DR WPI: 96-2211643/22.
DR P-PSDB: R94906.
PT New gene encoding a radiation protecting checkpoint protein - useful
PT for diagnosis and treatment of cancer and other diseases involving
PT abnormal apoptosis
PS Claim 2; Fig 3; 29pp; English.
CC A cDNA clone (T18696) codes for a novel radiation-protecting human
CC checkpoint protein (R94906), designated RAP-1, involved in
CC regulation of cell cycle progression and/or apoptosis. It was
CC isolated from a human cDNA library established in an immortalised
CC xeroderma pigmentosa cell line, GM2096 (XPIM1). A cDNA (cDNA3) contg.
CC a 462 bp insert complemented UV sensitivity and was used to screen a
CC lambda-gt10-X562 human cDNA library. The insert represented part of
CC the 3510 bp RAP-1 mRNA. The human checkpoint gene can be used as a
CC marker for early detection and prevention of tumour progression or
CC can be used to design new drugs that enhance radiotherapy or
CC chemotherapy of diseases involving abnormal apoptosis.
SQ Sequence 3510 BP; 974 A; 795 C; 848 G; 892 T;

Query Match 13.0%; Score 31; DB 1; Length 3510;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1 GGCACCGCGGACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCACTATGGGCTGTG 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1757 GCCACGCTGTGACCCGTCCTCCATGGGAGACCGAGAGAAATACATCTCTA 1816
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 61 GGCCTGCCCCCTGCTCTCTGTGACCTCTCTTGGCAGCTCTACATGAACAGGCGCGGTA 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1817 TCCTCTCTCTGGTATACCTCTCTTGGACTTCTCCAAAGAAACAGAAAGAGGAGGAT 1876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 121 TGACTTTGCAACTGAAGCTGAAG 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1877 CTAGTTGGCAGCTTAACGAGG 1899
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells
```

```
T02964/c
ID T02964 standard; cDNA; 580 BP.
AC T02964;
DE 01-MAR-1996 (first entry)
DE Cell-cycle regulatory protein p15 cDNA.
KW Cell-cycle regulatory protein p15; cyclin-dependent kinase inhibitor;
KW CCR; gene therapy; transgenic animal; cancer; cell proliferation;
KW ss; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT cds 91..483
FT FT /*tag= a
FT PN WO9528483-A1.
FT PD 26-OCT-1995.
FT PP 14-APR-1995; U04636.
FT PR 14-APR-1994; US-227371.
FT PR 25-MAY-1994; US-248812.
FT PR 14-SEP-1994; US-306511.
FT PR 29-NOV-1994; US-346147.
FT PA (COLD-) COLD SPRING HARBOR LAB.
FT PI Beach DH, Demetrick DJ, Hannon GJ, Serrano M;
FT WPI: 95-373798/48.
FT P-PSDB: R85118.
FT PT New cell cycle regulating proteins bind to cyclin dependent kinase -
FT and related nucleic acids, antibodies etc., used in diagnosis and
FT therapy of abnormal cell proliferation, degeneration etc.
FT PS Claim 43; Page 81-82; 109pp; English.
FT CC cDNA (T02964) coding for the mouse cell-cycle regulatory (CCR)
FT CC protein p15 (R85118) was isolated from an embryonal carcinoma
FT CC library using a probe based on a mouse CCR p13.5 cDNA (T02965).
FT CC The isolated cDNA can be used; to detect mutations in CCR
FT CC genes that lead to cell proliferation; to breed transgenic
FT CC animals to study cellular disorders involving CCR allele
FT CC mutation/misexpression; and to correct CCR-deficient cells
FT CC (gene therapy).
SQ Sequence 580 BP; 109 A; 187 C; 197 G; 87 T;

Query Match 12.7%; Score 30.4; DB 1; Length 580;
Best Local Similarity 57.3%; Pred. No. 1.3;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 39 GCAGTGGCCACTATGGGCTGTGGCTGGCCCTTGTCTCTCTTGTGACCTCTCTTGGCAGC 98
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 GCGGTGGCCAGGCCGCGTCACTGCTGCGCCCAACATGCGCTTGTCCCGGCTGTGTGC 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 99 TCACATGGAACAGGCGGGTATGACTTTTCAACTG 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AGAATGGTCTTCGCGCGCGGTGAGATTGCTACAG 33
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
V50430
ID V50430 standard; cDNA; 1301 BP.
AC V50430;
DE 07-DEC-1998 (first entry)
DE Mouse coxsackievirus and Ad2 and Ad5 receptor (MCAR) cDNA.
KW MCAR; coxsackievirus receptor; CVB; adenovirus; Ad2 receptor;
KW Ad5 receptor; mouse; infection; vaccine; gene therapy; ds.
OS Mus sp.
FH Key Location/Qualifiers
FT cds 157..1215
FT FT /*tag= a
FT PN WO9833819-A1.
FT PD 06-AUG-1998.
FT PP 30-JAN-1998; U01724.
FT PR 30-JAN-1997; US-036986.
FT PA (UYNY) UNIV NEW YORK STATE.
FT PI Philipson L, Tomko RP;
FT WPI: 98-437397/37.
FT P-PSDB: W69698.
PT DNA encoding human receptor for adenovirus C and coxsackievirus B -
PT for preventing and treating viral infection and rendering cells
```

PT susceptible to transformation by adenoviral vectors in gene therapy
 PS Claim 2; Page 68-70; 88pp; English.
 CC This cDNA molecule codes for mouse MCAR protein (see W69698) that
 CC serves as a cellular receptor for adenoviruses of the serotypes 2
 CC and 5 (subgroup C) and for the group B coxsackieviruses (CVB). It
 CC was obtained by screening a lambda phage expression cDNA library
 CC with antiserum containing antibodies specific for mouse CVB
 CC binding protein p46, and was used to identify human HCAR cDNA
 CC (see V50429). The invention also provides host cells transformed
 CC with DNA molecules encoding HCAR or MCAR and methods of producing
 CC the recombinant proteins or their derivatives. These proteins,
 CC their extracellular domains, as well as oligopeptides (see
 CC W69699-708) which bind virus, are also provided. Isolated HCAR or
 CC MCAR proteins or their fragments or variants are used to prevent
 CC or treat virus infections. Expressing the DNA in cells which lack
 CC these viral receptors renders the cells susceptible to
 CC transformation by adenoviral vectors carrying genes for gene
 CC therapy.
 SQ Sequence 1301 BP; 355 A; 327 C; 322 G; 297 T;
 Query Match 12.7%; Score 30.4; DB 1; Length 1301;
 Best Local Similarity 61.2%; Pred. No. 1.8;
 Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 33 AGGAGCGCAGTGGCCACTATGGGTGGTGGCCCTTGTCTCTCTGACCTCTCTT 92
 DB 868 ACGATCGCGCGCGCGCTCATAGGAGCGTGTGGCCCTTGTCTCTGACCTCTCT 927
 QY 93 GGCAGCTCACATGGGAACAGG 112
 DB 928 TTCTGTGTCTACAGGAAACG 947
 RESULT 7
 ID V86365/c V86365 standard; cDNA; 291 BP.
 AC V86365;
 DT 27-APR-1999 (first entry)
 DE EST clone AR34.
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; chromolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 OS Homo sapiens.
 PN WO9845435-A2.
 PD 15-OCT-1998.
 PF 10-APR-1998; U06954.
 PR 10-APR-1997; US-835913.
 PA (GENEY) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI; 99-070076/06.
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 210; 633pp; English.
 CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC activity. The EST sequences are also stated to be useful for gene
 CC therapy.
 SQ Sequence 291 BP; 102 A; 55 C; 57 G; 77 T;

Query Match 12.5%; Score 29.8; DB 1; Length 291;
 Best Local Similarity 70.2%; Pred. No. 1.6;
 Matches 40; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
 QY 119 TATGACTTGCACCTGAGCTGAGGAGTCTTTCTGACAAATTCCTCTATGATGTC 175
 DB 257 TTGATTTTGAACAGAGCTGAGATATTTTCTTACATTTTCTTAAAAATC 201
 RESULT 8
 ID Q05106 Q05106 standard; DNA; 2259 BP.
 AC Q05106;
 DT 02-NOV-1990 (first entry)
 DE Sequence encoding tumour-associated antigen GA733-1.
 DE Pancreatic carcinoma; GA733-1; cancer; ds.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 307..1275
 FT EP-376746-A.
 PN 4-JUL-1990.
 PD 29-DEC-1988; 313687.
 PR 29-DEC-1988; US-291583.
 PA (WIST-) Wistar Inst.
 PI Linnebach A;
 DR WPI; 90-203091/27.
 DR P-PSDB; R05711.
 PT Tumour-associated antigen, GA733-2 -
 PT Expressed in pancreatic carcinoma cells, used for tumor
 PT immunotherapy.
 PS Claim 2, Fig 2A; 12pp; English.
 CC The peptide is produced by tumor cells, and Abs raised to the
 CC peptide may be used in diagnosis and therapy of human tumours.
 SQ Sequence 2259 BP; 460 A; 648 C; 639 G; 512 T;
 Query Match 12.1%; Score 29; DB 1; Length 2259;
 Best Local Similarity 52.0%; Pred. No. 6;
 Matches 65; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 21 CYCTCCATCCAGAGCGGCGCTGCGCTATGGTCTGGCTGCGCTGCTGCTGCTGCT 80
 DB 1120 CGCTTACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1179
 QY 81 TTGACCCCTCTTGGCAGCTCATCATGGAACAGCGCGGCTGCTGCTGCTGCTGCTG 140
 DB 1180 ATGGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1239
 QY 141 AAGGA 145
 DB 1240 AAGGA 1244
 RESULT 9
 ID N60409 N60409 standard; DNA; 1110 BP.
 AC N60409;
 DT 01-JAN-1980 (first entry)
 DE Human apolipoprotein-E.
 KW Apolipoprotein-E; Hyperlipidemia; arteriosclerosis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 15..968
 FT A08547513-A.
 PN 24-APR-1986.
 PD 17-SEP-1985; 047513.
 PR 11-JUN-1985; JP-126989.
 PA (MITU) MITSUBISHI CHEM IND KK.
 DR WPI; 86-150217/24.
 DR P-PSDB; P60507.

PT New DNA sequence coding for human apolipoprotein-E - and
PT expression vectors and transformed cells contg. it
PS Disclosure; Fig 2; 45pp; English.
CC The encoded protein is used to treat subjects who are deficient in
CC apolipoprotein-E (or who produce abnormal forms of this molecule)
CC and therefore are likely to suffer from hyperlipidemia, resulting in
CC arteriosclerosis. It can also be used to raise antisera for
CC detecting the protein deficiency or production of abnormal forms.
SQ Sequence 1110 BP; 198 A; 354 C; 415 G; 143 T;

Query Match 12.1%; Score 29; DB 1; Length 1110;
Best Local Similarity 54.1%; Pred. No. 4.7;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 10 GACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCTATGGGTGGGCTGCCGCC 69
DB 579 GAGCGCGGCTCAGCGCCATCCGAGCGCTGGGGCCCTGGTGGACAGGCGCGGTG 638

QY 70 TTGTCCCTCTTGACCTCTTGGCAGCTCAGTGAACAGGCGCGG 118
DB 639 CGGCGCCGACGTGGGCTCCCTGGCGCGCAGCGCTACAGGAGCGG 687

RESULT 10
N50450
ID N50450 standard; DNA: 1110 BP.
AC N50450;
DT 09-JAN-1992 (first entry)
DE Sequence encoding human apolipoprotein E.
KW Hyperlipaemia; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 15..988
FT mat_peptide 69..965
FT /*tag= a
FT J60118189-A.
PN 25-JUN-1985.
PF 29-NOV-1983; 224980.
PD 29-NOV-1983; JP-224980.
PA (MITU) MITSUBISHI CHEM IND KK.
DR WPI: 85-188003/31.
DR P-PSDB: F51204.
PT DNA fragment - contg. DNA which codes human apolipoprotein E for
PT treatment of hyperlipaemia.
PS Claim 3; Page 484; 8pp; Japanese.
CC The sequence may be used to produce the apolipoprotein E, useful in the
CC treatment of hyperlipaemia.
SQ Sequence 1110 BP; 198 A; 355 C; 414 G; 143 T;

Query Match 12.1%; Score 29; DB 1; Length 1110;
Best Local Similarity 54.1%; Pred. No. 4.7;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 10 GACTTCAGTCTCTCCATCCAGGAGCGCAGTGGCTATGGGTGGGCTGCCGCC 69
DB 579 GAGCGCGGCTCAGCGCCATCCGAGCGCTGGGGCCCTGGTGGACAGGCGCGGTG 638

QY 70 TTGTCCCTCTTGACCTCTTGGCAGCTCAGTGAACAGGCGCGG 118
DB 639 CGGCGCCGACGTGGGCTCCCTGGCGCGCAGCGCTACAGGAGCGG 687

RESULT 11
Q57953/C
ID Q57953 standard; cDNA; 2095 BP.
AC Q57953;
DT 11-SEP-1994 (first entry)
DE Human beta1.6-N-acetylglucosaminyltransferase cDNA.
KW Beta1.6-N-acetylglucosaminyl-transferase;
KW acetylglucosaminotransferase; enzyme; lung carcinoma; tumor;

KW cancer diagnosis; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 156..2093
FT /*tag= a

PN EP-585109-A.
PD 02-MAR-1994.
PF 24-AUG-1993; 306718.
PR 24-AUG-1992; JP-245950.
PR 06-AUG-1993; JP-237118.
PA (SUNR) SUNTORY LTD.
PA (TANU) TANIGUCHI N.
PI Nishikawa A, Taniguchi N, Yamaguchi N;
DR WPI: 94-067582/09.
DR P-PSDB: R48975.
PT New beta 1,6-N-acetyl glucose aminyl transferase - useful in
PT sugar synthesis and for identification of cancer inhibitors, also
PT related DNA, vectors and transformed cells
PS Claim 2; Page 16; 30pp; English.
CC DNA encoding this enzyme may be introduced into cells for
CC expression. The resulting enzyme may be used to introduce,
CC homogeneously, beta(1-6) branch structures into oligosaccharides.
CC The enzyme may also be used to screen for specific inhibitors
CC (possibly cancer metastasis inhibitors).
SQ Sequence 2095 BP; 602 A; 481 C; 492 G; 520 T;

Query Match 12.1%; Score 29; DB 1; Length 2095;
Best Local Similarity 48.2%; Pred. No. 5.9;
Matches 80; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 67 CCCTGTCTCTCTTGCACCTCTGACGCTCAGTGCACATGACACAGCGCGGTGACTT 126
DB 487 CCATTGACAAAGATTGTCCACCTTGCACCTCCCACTTGCCTGCAAAATGTTA 428

QY 127 TGCAGTGAAGCTGAAGAGTCTTTTCTGACAAATTCCTCTATGAGTCCAGCTTCTGG 186
DB 427 TCTAATACACAGCAAGGGTTTCTTCAGATCATAGCTGTCATGACTCCAGCGTATGGC 368

QY 187 AATTGCTTGAANAATCTGCTCTCTCTCCATCTCCCTTCAGGAGCC 232
DB 367 CCAATCCACCAATCTCTTCTTCTGCGAGTCCCTTGATGATACC 322

RESULT 12
Q69099/C
ID Q69099 standard; cDNA; 597 BP.
AC Q69099;
DT 26-JAN-1995 (first entry)
DE ApoE4L protease cDNA.
KW ApoE4L; protease; enzyme; Alzheimer disease; diagnostic; therapeutic; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 1..595
FT /*tag= a
FT /label= ApoE4L protease

PN W09413798-A.
PD 23-JUN-1994.
PF 16-DEC-1993; E03581.
PR 16-DEC-1992; CA-085924.
PR 04-MAR-1993; US-291401.
PA (BERG) BERGMANN J E.
PA (PRED) PREDDIE R E.
PI Bergmann JE, Preddie RE;
DR WPI: 94-234212/28.
DR P-PSDB: R39841.
PT New proteinase esterase-like proteins - used to develop prods.
PT for the diagnosis and treatment of Alzheimer's disease and
PT related diseases
PS Claim 9; Page 38-39; 72pp; English.
CC The cDNA encodes ApoE4L, a protease catalyzing the formation of the
CC abnormal beta/A4 variant of beta-amyloid protein, which is used
CC to develop an inhibitor for the diagnosis and treatment of Alzheimer

PT. related conditions, hyperlipoproteinaemia or cardiovascular disease

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 21, 2000, 23:48:36 ; Search time 50.59 Seconds
(without alignments)
565.599 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GCCCAGCGGACTTCAGTGT.....CCCTTCAGGACGACGCGTCA 239

Scoring table: IDENTITY_NUC

Searched: 214294 seqs, 59861208 residues

Database : Issued_Patents_NA.*

Word size : 0

Number of hits that pass the threshold : 428588
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/PTC059_COMB.seq.*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	31	13.0	3509	3	US-08-817-436A-1	Sequence 1, Appl
2	30.4	12.7	7218	1	US-08-232-463-14	Sequence 14, Appl
3	30.4	12.7	580	4	US-08-627-610-7	Sequence 7, Appl
C 4	30.4	12.7	580	6	PCT-US95-04636-7	Sequence 7, Appl
C 5	29	12.1	1157	1	US-07-709-949-1	Sequence 1, Appl
C 6	29	12.1	2095	2	US-08-405-230-8	Sequence 8, Appl
C 7	29	12.1	2095	3	US-08-910-990-8	Sequence 8, Appl
C 8	29	12.1	2259	7	5185254-3	Patent No. 5185254
C 9	27.6	11.5	5102	2	US-08-494-168-1	Sequence 1, Appl
C 10	27.6	11.5	3088	2	US-08-418-444A-1	Sequence 1, Appl
C 11	27.6	11.5	1294	4	US-08-665-647-4	Sequence 4, Appl
C 12	27.2	11.4	2122	2	US-08-485-449-1	Sequence 1, Appl
C 13	27.2	11.4	3398	6	PCT-US95-08493-12	Sequence 12, Appl
C 14	27	11.3	1593	3	US-08-878-989-9	Sequence 9, Appl
C 15	26.8	11.2	1554	2	US-08-469-486-1	Sequence 1, Appl
C 16	26.8	11.2	3065	4	US-08-852-153-7	Sequence 7, Appl
C 17	26.8	11.2	1554	4	US-08-469-658-1	Sequence 1, Appl
C 18	26.8	11.2	2703	4	US-08-288-508C-1	Sequence 1, Appl
C 19	26.6	11.1	1241	3	US-08-169-948B-11	Sequence 11, Appl
C 20	26.6	11.1	1241	3	US-08-448-873-11	Sequence 11, Appl
C 21	26.6	11.1	1380	4	US-08-766-856A-4	Sequence 4, Appl
C 22	26.6	11.1	1300	4	US-09-166-203-41	Sequence 41, Appl
C 23	26.4	11.0	37895	1	US-08-375-709-1	Sequence 1, Appl
C 24	26.4	11.0	8268	1	US-08-375-709-10	Sequence 10, Appl
C 25	26.4	11.0	37895	2	US-08-752-929-1	Sequence 1, Appl
C 26	26.4	11.0	8268	2	US-08-752-929-10	Sequence 10, Appl
C 27	26	10.9	44377	3	US-08-804-237C-7	Sequence 7, Appl
C 28	26	10.9	44377	4	US-08-804-158-1	Sequence 1, Appl
C 29	25.8	10.8	2334	1	US-08-406-070-1	Sequence 1, Appl
C 30	25.8	10.8	655	3	US-08-812-645-2	Sequence 2, Appl
C 31	25.6	10.7	944	1	US-08-665-617-1	Sequence 1, Appl
C 32	25.6	10.7	1215	7	5169941-7	Patent No. 5169941

33	25.4	10.6	252	2	US-08-689-190-1	Sequence 1, Appl
34	25.4	10.6	263	2	US-08-689-190-3	Sequence 3, Appl
C 35	25.4	10.6	265	2	US-08-689-190-4	Sequence 4, Appl
C 36	25.4	10.6	234	3	US-08-733-446-19	Sequence 19, Appl
37	25.4	10.6	237	3	US-08-733-446-20	Sequence 20, Appl
38	25.4	10.6	240	3	US-08-733-446-21	Sequence 21, Appl
39	25.4	10.6	243	3	US-08-733-446-22	Sequence 22, Appl
40	25.4	10.6	243	3	US-08-733-446-23	Sequence 23, Appl
41	25.4	10.6	245	3	US-08-733-446-24	Sequence 24, Appl
C 42	25.4	10.6	247	3	US-08-733-446-25	Sequence 25, Appl
C 43	25.4	10.6	248	3	US-08-733-446-26	Sequence 26, Appl
C 44	25.4	10.6	250	3	US-08-733-446-27	Sequence 27, Appl
45	25.4	10.6	251	3	US-08-733-446-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-08-817-436A-1
; Sequence 1, Application US/08817436A
; Patent No. 5882880
; GENERAL INFORMATION:
; APPLICANT: Canaani, Dan
; TITLE OF INVENTION: Human Checkpoint Gene and Gene for
; TITLE OF INVENTION: Antisense RNA thereof
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wigman, Cohen, Leitner & Myers
; STREET: 900 17th Street, N.W., Suite 1000
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.5 inch, 1.44Mb storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,436A
; FILING DATE: 11-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/12445
; FILING DATE: 11-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cohen, Herbert
; REGISTRATION NUMBER: 25,109
; REFERENCE/DOCKET NUMBER: 0744.066
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-463-7700
; TELEFAX: 202-463-6915
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3509 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL LINE: ESTABLISHED XERODERMA PIGMENTOSUM GM2096-SV3
; IMMEDIATE SOURCE:
; CLONE: RAP-1 cDNA
US-08-817-436A-1

Query Match 13.0%; Score 31; DB 3; Length 3509;
Best Local Similarity 51.0%; Pred. NO. 0.45;
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

PCT-US95-04636-7/c
; Sequence 7, Application PC/TUS9504636
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins, and Uses
; NUMBER OF SEQUENCES: 10
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04636
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/346,147
; FILING DATE: 29-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/306,511
; FILING DATE: 14-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/248,812
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/227,371
; FILING DATE: 14-APR-1994
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 91..480
PCT-US95-04636-7

Query Match 12.7%; Score 30.4; DB 6; Length 580;
Best Local Similarity 57.3%; Pred. No. 0.31;
Matches 55; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
QY 39 GCATGGCCACTATGGGTGGGCTGCGCCCTTGTCTCTTGACCTCTCTGGCAGC 98
Db 128 GCGTGGCCAGCGCCGCGTCACTGCTGCGCCCAACATGCGCCCTGTCCCGGCTGTGGC 69
QY 99 TCACATGGACAGCGCGGGGTATGACTTTGCAACTG 134
Db 68 AGAATGGTCTTCGCGCGCGGTGAGATGCTACAG 33

RESULT 5
US-07-709-949-1
; Sequence 1, Application US/07709949
; Patent No. 5472858
; GENERAL INFORMATION:
; APPLICANT: Attie, Alan D
; APPLICANT: Gretch, Daniel G
; APPLICANT: Sturley, Stephen L
; APPLICANT: Beckage, Nancy E
; TITLE OF INVENTION: Production of Recombinant Proteins in
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: P.O. Box 2113
; CITY: Madison
; STATE: WI
; COUNTRY: USA
; ZIP: 53701-2113

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/709,949
; FILING DATE: 19910604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 9629691801
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1157 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 62..1015
; FEATURE:
; NAME/KEY: misc_signal
; LOCATION: 62..119
; PUBLICATION INFORMATION:
; AUTHORS: McLean, J W
; JOURNAL: Journal of Biological Chemistry
; VOLUME: 259
; PAGES: 6498-6504
; DATE: 1984
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 1157
US-07-709-949-1

Query Match 12.1%; Score 29; DB 1; Length 1157;
Best Local Similarity 54.1%; Pred. No. 1.3;
Matches 59; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 10 GACTTCAGTGTCTCTCCATCCAGGAGCGCAGTGGCCACTATGGGCTCTGGGCTGCCCC 69
Db 626 GAGCGGCGCTCAGCGCCATCCGAGCGCTGGGGCCCCCTGGTGGACAGGCGCGGTG 685
QY 70 TTGCTCTCTCTTGACCTCTCTTGGCAGCTCACAATGGAACAGGCGCGG 118
Db 686 CGGGCGGCGCACTGTGGGCTCCCTGCGCGGCGCAGCGCTACAGGAGCGGG 734

RESULT 6
US-08-405-230-8/c
; Sequence 8, Application US/08405230
; Patent No. 5707846
; GENERAL INFORMATION:
; APPLICANT: TANIGUCHI, Naoyuki
; APPLICANT: NISHIKAWA, Atsushi
; APPLICANT: YAMAGUCHI, No. 5707846om1
; TITLE OF INVENTION: NOVEL N-ACETYLGLUCOSAMINYL TRANSFERASE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: George Mason Bldg., Washington & Prince Sts.
; CITY: Alexandria
; STATE: Virginia

COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,230
FILING DATE: 16-MAR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/110,736
FILING DATE: 23-AUG-1993
APPLICATION NUMBER: JP 4-245950
FILING DATE: 24-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-237118
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 001560-215
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2095 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 156..2093
US-08-405-230-8

Query Match 12.1%; Score 29; DB 2; Length 2095;
Best Local Similarity 48.2%; Pred. No. 1.6;
Matches 80; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 67 CCCTTGCTCCTCTTGACCCCTCTTGGCAGCTCAGATGGAACAGGCGCGGTATGACTT 126
DB 487 CCATGTGACAAAGATTGTCCACCTTCGACTCCCAATGGCGTGCAGAAATGTTA 428
QY 127 TGAACCTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCTATGATGATCCAGCTTCTGG 186
DB 427 TCTAATAACACAGCAGGAGTTTCTTCAGATCATAGCTGTGATGCTCCAGCGTATGCC 368
QY 187 AATTGCTTGAAGTCTGCTCCCTCTCCCTCCATCTCCCTTCAGGGACC 232
DB 367 CCATCCACCACATTCCTGTTTCTTCTGCGAGTGCCTTGATGATACC 322

RESULT 7
US-08-910-990-8/c
Sequence 8, Application US/08510990
Patent No. 5834284
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: NISHIKAWA, Atsushi
APPLICANT: YAMAGUCHI, No. 5834284om1
TITLE OF INVENTION: NOVEL N-ACETYLGLUCOSAMINYL TRANSFERASE
TITLE OF INVENTION: GENE CODING THEREFOR AND PROCESS FOR PRODUCTION THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
CITY: Alexandria
STATE: Virginia
COUNTRY: United States

ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,990
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,230
FILING DATE: 16-MAR-1995
APPLICATION NUMBER: US 08/110,736
FILING DATE: 23-AUG-1993
APPLICATION NUMBER: JP 4-245950
FILING DATE: 24-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 5-237118
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Crane-Feury, Sharon E
REGISTRATION NUMBER: 36,113
REFERENCE/DOCKET NUMBER: 001560-215
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2095 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 156..2093
US-08-910-990-8

Query Match 12.1%; Score 29; DB 3; Length 2095;
Best Local Similarity 48.2%; Pred. No. 1.6;
Matches 80; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 67 CCCTTGCTCCTCTTGACCCCTCTTGGCAGCTCAGATGGAACAGGCGCGGTATGACTT 126
DB 487 CCATGTGACAAAGATTGTCCACCTTCGACTCCCAATGGCGTGCAGAAATGTTA 428
QY 127 TGAACCTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTCTATGATGATCCAGCTTCTGG 186
DB 427 TCTAATAACACAGCAGGAGTTTCTTCAGATCATAGCTGTGATGCTCCAGCGTATGCC 368
QY 187 AATTGCTTGAAGTCTGCTCCCTCTCCCTCCATCTCCCTTCAGGGACC 232
DB 367 CCATCCACCACATTCCTGTTTCTTCTGCGAGTGCCTTGATGATACC 322

RESULT 8
5185254-3
Patent No. 5185254
APPLICANT: LINNENBACH, ALBAN
TITLE OF INVENTION: GENE FAMILY OF TUMOR-ASSOCIATED ANTIGENS
NUMBER OF SEQUENCES: 5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/291,583
FILING DATE: 29-DEC-1988
SEQ ID NO: 3
LENGTH: 2259
5185254-3
Query Match 12.1%; Score 29; DB 7; Length 2259;
Best Local Similarity 52.0%; Pred. No. 1.7;


```

, , TELEPHONE: 415-955-0555
, ,
, , TELEFAX: 415-945-4166
, ,
, , TELEX:
, ,
, , INFORMATION FOR SEQ ID NO: 9:
, ,
, , SEQUENCE CHARACTERISTICS:
, , LENGTH: 1593 base pairs
, , TYPE: nucleic acid
, , STRANDEDNESS: single
, , TOPOLOGY: linear
, , IMMEDIATE SOURCE:
, , LIBRARY: TBLYN01
, , CLONE: 40194
, ,
, , US-08-078-989-9

```

	Query Match	11.38;	Score 27;	DB 3,	Length 1593;
	Best Local Similarity	52.8%;	Pred. No. 6.8;	Indels	0;
	Matches	57;	Conservative	0;	Caps
120	ATGACTTTGC	AACTGA	CTGAGGAGCTTTTCTTG	CAAAANTCTCCT	TATGATGCCACG 179
1168	ATGACTTTGG	GAATTCG	TATAGCTGATTTTTT	TGGGAAACTGTTT	TACTTCATTTCAAG 1109
180	TTCCTGGAA	TGCTT	GAAANNTCGCTCCT	CTCCATCTCCCTTCAG 227	
1108	GTTCCTGAG	ACTCTG	TATATTCCTGCTCT	CTCTGCTTCCTCTGAG 1061	

```

15
US-08-469-486-1
; Sequence 1, Application US/08469486
; Patent No. 5739281
; GENERAL INFORMATION:
; APPLICANT: Thøgersen, Hans Christian
; APPLICANT: Holtz, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: Improved method for the refolding of
; TITLE OF INVENTION: proteins
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,486
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1554 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1551
US-08-469-486-1

Query Match          11.2%; Score 26.8; DB 2; Length 1554;
Best Local Similarity 49.3%; Pred. No. 7.8;
Matches 70; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 46 CCACATATGGGGTCTGGGGCTGGCCCTCTGTCTCTCTTGACCTCTCTGTGGCAGCTACATG 105
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 CCACCATATGGGGGCTCTGTCATCTCTGCTTGTCTGACGACCGCCTCTGGCGGCGCTCTGC 130
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 106 GAACAGCGCGCGGTATGACTTTTGCAACTGAAGCTGAAGAGGTCTTTTCTGCAAAATTCCT 165
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 GGCCGGCGGGGAGCGCTTCTCTGCGCCGGGACGAGCCGCCACCGTCTCTGACAGAGCCG 190
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 166 CCTATGAGTCCAGCTTCCTTGGG 187
      ||||| ||||| ||||| |||||
Db 191 GCAGGGCCAACTCATTCCTTGGG 212
      ||||| ||||| ||||| |||||

```

Search completed: March 22, 2000, 01:22:29
Job time: 5633 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: March 22, 2000, 04:46:59 ; Search time 802.03 Seconds
(without alignments)
1125.125 Million cell updates/sec

Title: US-09-092-296-1
Perfect score: 239
Sequence: 1 GGCACCGGAGCTTCAGTGT.....CCCTCAGGAGCACCAGCTCA 239
Scoring table: IDENTITY_NUC
Searched: 4538634 seqs, 1887831982 residues
Database: EST.*

Word size : 0
Number of hits that pass the threshold : 9077268

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: em_est20:*
21: em_est21:*
22: em_est22:*
23: em_est23:*
24: em_est24:*
25: em_est25:*
26: em_est26:*
27: em_est27:*
28: em_est28:*
29: em_est29:*
30: em_est30:*
31: em_est31:*
32: em_est32:*
33: em_est33:*
34: em_est34:*
35: em_est35:*
36: em_est36:*
37: em_est37:*
38: em_est38:*
39: em_est39:*
40: em_est40:*
41: em_est41:*
42: em_est42:*
43: em_est43:*
44: em_est44:*
45: em_est45:*
46: em_est46:*
47: em_est47:*
48: em_est48:*
49: em_est49:*

50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_gss1:*
80: gb_gss2:*
81: gb_gss3:*
82: gb_gss4:*
83: em_gss1:*
84: em_gss2:*
85: em_gss3:*
86: em_gss4:*
87: gb_gss5:*
88: gb_gss6:*
89: gb_gss7:*
90: gb_gss8:*
91: gb_gss9:*
92: em_gss5:*
93: em_gss6:*
94: em_gss7:*
95: em_gss8:*
96: em_gss9:*
97: em_gss10:*
98: em_gss11:*
99: gb_gss10:*
100: gb_gss11:*
101: em_gss12:*
102: gb_gss12:*
103: gb_gss13:*
104: gb_gss14:*
105: gb_gss15:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	118	49.4	404	61	AI857998
C 2	74.4	31.1	552	82	AQ718761
C 3	63.6	26.6	328	42	AI136523
C 4	35.6	14.9	509	88	AQ838514
C 5	35	14.6	348	21	F06958
C 6	34.4	14.4	238	32	AA376266
C 7	34.4	14.4	412	38	AA769782

AI857998 wj69b01.x
AQ718761 HS_5511_B
AI136523 UI-R-C2p-
AQ838514 HS_5011_A
F06958 HSG1QC101 n
AA376266 EST88915
AA769782 eh71b05.s

c 8 34.4 14.4 458 49 A1632159
 c 9 34.4 14.4 470 60 A1803529
 c 10 32.8 13.7 513 60 A1807980
 c 11 32.6 13.6 360 63 AW009894
 c 12 32.4 13.6 537 64 AW078074
 c 13 32.4 13.6 523 69 AW128453
 c 14 32.2 13.5 339 21 F08745
 c 15 32.2 13.5 742 27 Z78408
 c 16 32.2 13.5 477 45 A1351147
 c 17 32.2 13.5 318 74 AW218386
 c 18 31.8 13.3 568 100 AQ290553
 c 19 31.6 13.2 382 36 AA620516
 c 20 31.6 13.2 438 38 AA797411
 c 21 31.6 13.2 427 59 A1776288
 c 22 31.4 13.1 505 20 M89011
 c 23 31.2 13.1 410 42 A1094732
 c 24 31.2 13.1 859 79 CNS000ED
 c 25 31.2 13.1 537 99 A178189
 c 26 31 13.0 535 105 AQ625402
 c 27 30.8 12.9 368 23 H28604
 c 28 30.8 12.9 589 51 A1730494
 c 29 30.8 12.9 313 61 A1810515
 c 30 30.6 12.8 573 23 T41535
 c 31 30.6 12.8 527 45 A1360929
 c 32 30.6 12.8 308 45 A1394138
 c 33 30.6 12.8 667 48 A1553740
 c 34 30.6 12.8 709 61 A1866474
 c 35 30.4 12.7 392 26 W70374
 c 36 30.4 12.7 290 33 AA387480
 c 37 30.4 12.7 407 74 AV395857
 c 38 30.4 12.7 445 87 A0768084
 c 39 30.2 12.6 383 40 AA934009
 c 40 30.2 12.6 530 62 A1881867
 c 41 30.2 12.6 603 64 AW066022
 c 42 30.2 12.6 455 64 AW066263
 c 43 30.2 12.6 883 79 CNS00CVS
 c 44 30.2 12.6 446 90 AQ068076
 c 45 30.2 12.6 461 99 AQ224224

ALIGNMENTS

RESULT 1
 A1857998/c
 LOCUS wj69b01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408041 3',
 DEFINITION mRNA sequence.
 ACCESSION A1857998
 VERSION A1857998.1 GI:5511614
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 404)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189004.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 cDNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 395.
 Location/Qualifiers
 1. 404
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:2408041"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 97 a 105 c 117 g 84 t 1 others
 ORIGIN
 Query Match 49.4%; Score 118; DB 61; Length 404;
 Best Local Similarity 97.7%; Pred. No. 7.3e-25;
 Matches 129; Conservative 0; Mismatches 2; Indels 1;
 Gaps 1;
 QY 109 CAGGCGCGGTGACTTGCACCTGAAGCTGAAGTCTTTCTGACAAATCTCTCTCT 168
 Db 311 CAGGCGCGGTGACTTGCACCTGAAGCTGAAGTCTTTCTGACAAATCTCTCTCT 252
 QY 169 ATGAGTCCAGTTCCTGGAATGCTTGAATAA-NTCTGCTCTCTCTCTCTCTCTCAG 227
 Db 251 ATGAGTCCAGTTCCTGGAATGCTTGAATAAAGCTNTGCTCTCTCTCTCTCTCTCAG 192
 QY 228 GGACGACGCTCA 239
 Db 191 GGACGACGCTCA 180

RESULT 2
 A0718761/c
 LOCUS A0718761 552 bp DNA GSS 13-JUL-1999
 DEFINITION HS_5511_B2_F09_77A RPCI-11 Human Male BAC Library Homo sapiens
 genomic clone Plate-1087 Col=18 Row=L, genomic survey sequence.
 ACCESSION A0718761
 VERSION A0718761.1 GI:5468077
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 552)
 AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong
 (pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (Info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
 Plate: 1087 row: L column: 18
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 552.

FEATURES
 source
 1..552
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate-1087 Col-18 Row-L"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACE3.6; Genomic sequence of BAC ends"

BASE COUNT 141 a 153 c 124 g 117 t 17 others
 ORIGIN

Query Match 31.1%; Score 74.4; DB 82; Length 552;
 Best Local Similarity 87.1%; Pred. No. 6.3e-12;
 Matches 81; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 107 AACAGGCGGGGATGACCTTCAACTGAAGCTGAAGGAGTCTTTCTGACAAATTCCTC 166
 Db 474 AGCATGGCGAGGATGACCTTCANCTGAAGCTGAAGGAGTCTATTAGACAAATTCGTC 415
 Qy 167 CTATGATCCAGCTTCTCGAATTCCTTGAAAA 199
 Db 414 TTATGATCCAGCTTCTCGAATTCCTTGAAAA 382

RESULT 3
 AII136523/C
 LOCUS
 DEFINITION UI-R-C2p-nq-e-02-0-UI-s1 UI-R-C2p Rattus norvegicus cDNA clone
 UI-R-C2p-nq-e-02-0-UI-3', mRNA sequence.
 ACCESSION AII136523
 VERSION
 KEYWORDS EST.
 SOURCE AII136523.1 GI:3637300
 ORGANISM Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 328)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 9704477
 Genome Res. 6 (9), 791-806 (1996)
 On Jan 19, 1998 this sequence version replaced gi:2150222.
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..328
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="UI-R-C2p-nq-e-02-0-UI"
 /clone_stage="adult"
 /lab_host="DH10B (Life Technologies)"

FEATURES
 source
 1..328
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone_lib="UI-R-C2p-nq-e-02-0-UI"
 /clone_stage="adult"
 /lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-C2p library is a subtracted library derived from the UI-R-C1 library, which is a subtracted library derived from the UI-R-C0 library. The UI-R-C0 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C2p) was constructed as follows: PCR amplified cDNA inserts from UI-R-C1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C1 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C2p library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)".

BASE COUNT 62 a 77 c 98 g 91 t
 ORIGIN

Query Match 26.6%; Score 63.6; DB 42; Length 328;
 Best Local Similarity 68.7%; Pred. No. 7.9e-09;
 Matches 101; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

Qy 94 GCAGCTCACATGACAGCGCGGTATGACTTTCGAACCTGAAGGAGTCTTTTC 153
 Db 327 GCAGCTCACATGACAGCGCGGTATGACTTTCGAACCTGAAGGAGTCTTTTC 268
 Qy 154 TGACAAATTCCTCTATGAGTCCAGTCTCTGGAATTCCTGAAAA-NTCGCTCCCTCC 212
 Db 267 AGGCAAGACCTCCCAAGACTCCGGCTTCTGACATGCTCCAAAAGATCTGCTCCCTCC 208
 Qy 213 TCCATCTCCCTTCAGGACGACGCTCA 239
 Db 207 TCCACCTCTCAGCGGGACCAATGCTCA 181

RESULT 4
 A0838514
 LOCUS
 DEFINITION HS.5011.A2.D10.SP6E RPCI-11 Human Male BAC Library Homo sapiens
 genomic Clone Plate-587 Col-20 Row-G, genomic survey sequence.
 ACCESSION A0838514
 VERSION
 KEYWORDS GSS.
 SOURCE A0838514.1 GI:5808388
 ORGANISM human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 509)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kaller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (<http://www.htsc.washington.edu>)
 Plate: 587 row: G column: 20
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 509.

FEATURES
 source
 1. .509
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="RPMI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Genomic sequence of BAC ends"

BASE COUNT 120 a 127 c 106 g 153 t 3 others
 ORIGIN
 Query Match 14.9%; Score 35.6; DB 88; Length 509;
 Best Local Similarity 47.3%; Pred. No. 1.8;
 Matches 104; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 8 GGGACTTCAGTGTCTCTCCATCCAGGAGCGAGTGGCCATATGGGGTCTGGCTGCC 67
 Db 80 GTGACCTCTTTCTCTCACCCCATGTGCTCAGGAGGAGAGCAGGTGGCTTTCTTC 139
 Qy 68 CTTGTCTCTCTCTTGGACCTCTTGGCAGCTCACATGGAACAGGCGCGGTATGACTTT 127
 Db 140 CCTCTTGCTAATATGCTCTCCCTGGCAAGCTGCCAACCACTGGTCTCTTACAGCTA 199
 Qy 128 GCACTGAGCTGAGGAGTCTTTCTGACAAATCTCTATGAGTCCAGCTCTCTGGA 187
 Db 200 GCTAGTGGAGATGGCCAGGTGATGTATAAATGAGCCTCTGCTGTGAGTCCAGGCC 259
 Qy 188 ATTGCTTGAANAATCTCCCTCTCTCTCCATCTCCCTTCAG 227
 Db 260 ATGTAAGTTACAGTGGCACCAGCACTGGGCTCTGGAG 299

RESULT 5
 LOCUS F06958 348 bp mRNA EST 20-FEB-1995
 DEFINITION HSCIQC101 normalized infant brain cDNA Homo sapiens cDNA clone c-lqc10, mRNA sequence.

ACCESSION F06958
 VERSION F06958.1 GI:672595
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 348)
 AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Juneau, M.N., Lamy, B., Lorenzo, F., Michell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y., Sebastiani-Kabaktchis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome and its expression
 JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
 MEDLINE 95277534
 COMMENT On Sep 21, 1992 this sequence version replaced gi:278976.
 Contact: Genethon
 Genexpress-Genethon
 Genethon Centre de recherche sur le Genome Humain
 1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
 Tel: 33169472800
 Fax: 33160778698
 Email: genexpress@genethon.fr
 Single read.
 Genexpress_library_id: C; Genexpress_sequence_id: yic-lqc10
 Insert Length: 639 Std Error: 0.00

Seq primer: (-21)M13_universal
 High quality sequence stop: 150.
 Location/Qualifiers
 1. .348
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="normalized infant brain cDNA"
 /sex="Female"
 /tissue_type="total brain"
 /dev_stage="3 months old"
 /note="Organ: brain; Vector: lafmid BA; Site 1: HindIII; Site 2: NotI; sex:Female; dev_stage=3 months old; isolate-muscular atrophy patient; tissue_type=total brain; total mRNA was oligo-(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lafmid BA vector. Clone library from B.Soures, Psychiatry Dept. Columbia University, USA. Normalization_method: Bento Soares, P.N.A.S in press"

BASE COUNT 84 a 91 c 81 t 2 others
 ORIGIN

Query Match 14.68; Score 35; DB 21; Length 348;
 Best Local Similarity 63.1%; Pred. No. 2.3;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 Qy 18 TGTCTCTCCATCCAGGAGCGAGTGGCTGTGGTGGTGGCTGTGGCTGTGGCTGTGGCT 77
 Db 5 TGTGGCTCTCTCCAGGATCCCTTGTGTGATGTGTTCAGGNTGCACCACCAC 64
 Qy 78 CTCCTGACCTCTTGGCAGCTCA 101
 Db 65 CTCAGATACCTTCAGGCAACACA 88

RESULT 6
 LOCUS AA376266 238 bp mRNA EST 21-APR-1997
 DEFINITION EST88915 HSCI72 cells II Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA376266
 VERSION AA376266.1 GI:2028809
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 238)
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M., Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Peligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, F.A., Collins, E.J., Dunke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 12140200
 COMMENT On Apr 14, 1993 this sequence version replaced gi:693135.
 Other ESTs: THCI91210
 Contact: Kerlavage, AR
 Bioinformatics

	BASE COUNT	108 a	56 c	78 g	108 t	
Query Match						
Best Local Similarity						
Matches						
Conservative						
Score 32.6;						
DB 63;						
Length 360;						
Pred. No. 12;						
0; Mismatches 100;						
Indels 0; Gaps						
0;						
a	28	ATCCAGAGCGCATGCGCCACTATGCGGTCTGGGTGCGCCCTTGCTCTCTCTTGACCC	87			
b	57	ATACAAGATTGAGTGCTCCACGCTTACCTTAAGCTGTATCTGTATTAGGATTGAATC	116			
c	88	TCCTTGGCAGCTCATGTCAGACAGCGCGGGTATGACTTTGCACTGAAGCTGAAGAGAGT	147			
d	117	TTCTTGAGTGGGCTGTGTAACCTGGTAAGATPATACTGCTAGGACTCAACTTGAAGAGA	176			
e	148	CTTTTCTGACAAATTCCTCTATGAGTCCAGCTTCTTGGAAATTCGTTGAAAATTCGCT	207			
f	177	ATTTGAGGAGACACCCGCCACATGCAGATGGGACCAAGAGTGGCTGGRACATGATGCA	236			
g	208	CTCTCTCCATCT	219			

BASE COUNTY
ORIGIN

I/3 a	134 c	120 g	109 t	1 others
-------	-------	-------	-------	----------

Dept. Columbia University, USA. Normalization_method:

Bento Soares, P.N.A.S in press*

BASE COUNT 57 a 116 c 100 g 64 t 2 others

ORIGIN

Query Match 13.5%; Score 32.2; DB 21; Length 339;

Best Local Similarity 58.5%; Pred. No. 15; Mismatches 39; Indels 0; Gaps 0;

Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 23 CTTCCATCCAGGAGCGAGTGGCCACTATGGGTCTGGGCTGCCCTTGTCTCTCTT 82

DB 61 CCCACTGGCGTGTCTCTCCGCCCTCCGAGCTTGGATGCCCTTCACACCCCTCTT 120

QY 83 GACCTCTCTTGCAGCTCACATGGACAGGCGC 116

DB 121 GATCTTCCCTGTGATGTACCTGGACCCCTGCTG 154

RESULT 15

278408

LOCUS 742 bp mRNA EST 28-JUL-1999

DEFINITION HS278408 Human fetal brain S. Meier-Ewert Homo sapiens cDNA clone

3.99 (CEPH), mRNA sequence.

ACCESSION 278408

VERSION 278408.1 GI:1495181

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 742)

AUTHORS

Neri, C., Albanese, V., Lebre, A.S., Holbert, S., Saada, C.,

Bougueret, L., Meier-Ewert, S., Legall, I., Millasseau, P., Bui, H.,

Giudicelli, C., Massart, C., Guillou, S., Gervy, P., Poullier, E.,

Rigault, P., Weissenbach, J., Lennon, G., Chumakov, I., Dausset, J.,

Lehrach, H., Cohen, D. and Cann, H.M.

Survey of CAG/CTG repeats in human cDNAs representing new genes:

candidates for inherited neurological disorders

Hum. Mol. Genet. 5 (7), 1001-1009 (1996)

96414310

On Apr 14, 1993 this sequence version replaced gi:837709.

Contact: Neri C.

Fondation Jean Dausset - CEPH

27 Rue Juliette Dodu, 75010 Paris, France

ICRF clone ID ICRF507L04199.

Location/Qualifiers

1. .742

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="3.99 (CEPH)"

/clone_lib="Human fetal brain S. Meier-Ewert"

/dev_stage="fetus"

/note="cDNA library of S. Meier-Ewert, Max Planck

Inst.f.Mol.Genetics, Berlin, FRG"

BASE COUNT 145 a 232 c 215 g 131 t 19 others

ORIGIN

Query Match

Best Local Similarity 13.5%; Score 32.2; DB 27; Length 742;

Matches 64; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 19 GTCTCTCTATCCAGGAGCGAGTGGCCACTATGGGTCTGGGCTGCCCTTGTCTCTCC 78

DB 198 GGCTACTCTAGGGATAGCCCGGCTCCCTCATCCCTGGTCAGTGTCTTGACCCCA 257

QY 79 TCTTGACCTCTCTTGGAGCTCACATGGAAACAGGCGGGGTATGACTTTGCACTGAAG 137

DB 258 TCTACTCTCTCTGGGGACTTCTCAGTCACAGGCTGGGCTGGGCTTTCANCTGNAG 316

Search completed: March 22, 2000, 05:09:43

Job time: 1364 sec

WIPERLH (TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 20 12:32:52 2000; MasPar time 5.78 Seconds
319.798 Million cell updates/sec
Tabular output not generated.

Title: >US-092-296-15
Description: (1-78) from US0902296.pep
Perfect Score: 558
Sequence: 1 MGSGPLVLLTLGSSHGT.....SGTSVTLHARSQHVVNCVT 78

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0\$
Listing first 45 summaries

Database: a-geneseq35
l-geneseq

Statistics: Mean 27.552; Variance 131.690; scale 0.209

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	558	100.0	78	Human stomach carcinoma	3.43e-38
2	86	15.4	46	Secreted protein FB78_	4.56e-01
3	86	15.4	47	Human secreted protein	4.56e-01
4	85	15.2	325	G-protein coupled bovi	5.33e-01
5	85	15.2	325	G-protein coupled bovi	5.33e-01
6	84	15.1	154	H. pylori GHPO 895 pro	6.24e-01
7	84	15.1	154	H. pylori GHPO 895 pro	6.24e-01
8	83	14.9	194	Peptide fragment of te	7.29e-01
9	83	14.9	411	Mouse apoptosis induci	7.29e-01
10	82	14.7	110	Human T lymphocyte rec	8.52e-01
11	82	14.7	169	Human mitochondrial el	8.52e-01
12	81	14.5	401	Human 7-transmembrane	9.94e-01
13	81	14.5	1253	Mad binding protein, m	9.94e-01
14	81	14.5	1261	Mad binding protein, m	9.94e-01
15	80	14.3	398	Human serotonin ID rec	1.16e+02
16	80	14.3	713	Cyclomaltodextrin gluc	1.16e+02
17	80	14.3	713	Sequence of cyclomalto	1.16e+02
18	80	14.3	945	Human NMDAR2 receptor	1.16e+02
19	80	14.3	1214	Human NMDAR2 receptor	1.16e+02
20	80	14.3	1214	Human N-methyl-D-aspar	1.16e+02
21	80	14.3	1219	Human N-methyl-D-aspar	1.16e+02
22	80	14.3	1219	Human NMDAR2 receptor	1.16e+02
23	80	14.3	1231	Human NMDAR2 receptor	1.16e+02

24	80	14.3	1231	1	W87506	Human N-methyl-D-aspar	1.16e+02
25	80	14.3	1236	1	R66037	Human N-methyl-D-aspar	1.16e+02
26	80	14.3	1236	1	W85574	Human N-methyl-D-aspar	1.16e+02
27	80	14.3	1239	1	W87508	Human N-methyl-D-aspar	1.16e+02
28	80	14.3	1239	1	R66064	Human NMDAR2 receptor	1.16e+02
29	80	14.3	1244	1	R66061	Human NMDAR2 receptor	1.16e+02
30	80	14.3	1244	1	W87505	Human N-methyl-D-aspar	1.16e+02
31	79	14.2	620	1	W14993	Human c-fos induced gr	1.35e+02
32	79	14.2	652	1	W49879	Amino acid sequence of	1.35e+02
33	79	14.2	907	1	W93889	Human HG38 protein.	1.35e+02
34	79	14.2	969	1	W25170	Human insulinoma-asso	1.35e+02
35	79	14.2	986	1	W25171	Human insulinoma-asso	1.35e+02
36	79	14.2	1015	1	W18092	Type I diabetes-associ	1.35e+02
37	79	14.2	1015	1	W35345	Human protein tyrosine	1.35e+02
38	78	14.0	99	1	W67838	Human secreted protein	1.58e+02
39	78	14.0	120	1	R37838	Kaposi's sarcoma assoc	1.58e+02
40	78	14.0	120	1	R93614	Kaposi's sarcoma assoc	1.58e+02
41	78	14.0	727	1	W11719	C-Delta-1 polypeptide.	1.58e+02
42	78	14.0	740	1	W08876	C-Delta-1 polypeptide	1.58e+02
43	78	14.0	922	1	W88419	Chlamydia pneumoniae s	1.58e+02
44	78	14.0	1239	1	R45945	Glutamic acid receptor	1.58e+02
45	77	13.8	246	1	W98638	H. pylori GHPO 343 pro	1.83e+02

ALIGNMENTS

RESULT 1
ID W88498 standard; Protein; 78 AA.
AC W88498;
DE 30-MAR-1999 (first entry)
DE Human stomach carcinoma HP10408-encoded transmembrane protein.
KW Transmembrane protein; HP10408; human; stomach cancer.
OS Homo sapiens.
PN W09855508-A2.
PD 10-DEC-1998.
PF 03-JUN-1998; J02445.
PR 03-JUN-1997; JP-144948.
PA (PROT-) PROTEGENE INC.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Kato S, Sekine S, Yamaguchi T;
DR WPI: 99-045730/04.
DR N-PSDB: W84366.
PT New human proteins containing transmembrane domains and their
PT encoding sequences - useful in the preparation of antibodies and
PT large-scale protein production, gene diagnosis, and gene therapy
PS Claim 1; Page 135; 178pp; English.
CC This is the amino acid sequence of a novel transmembrane protein
CC encoded by human stomach cancer cDNA clone HP10408 (see W84366).
CC The encoded protein has a putative signal sequence and a putative
CC internal transmembrane domain. The invention provides nucleotide
CC sequences (see W84359-76) coding for 18 transmembrane proteins
CC (see W88491-509), vectors containing such polynucleotides, and
CC eukaryotic cells containing the vectors. The proteins can be
CC used as antigens or as compositions in the preparation of
CC antibodies against the proteins. The polynucleotides can be used
CC as probes for gene diagnosis, and as gene sources for gene therapy
CC and large-scale production of proteins encoded by the cDNA. The
CC host cells are used for the detection of ligands corresponding to
CC the expressed proteins, and the screening of low mol.wt. medicines.
SQ Sequence 78 AA;

Query Match 100.0%; Score 558; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 3.43e-38;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	MGSGPLVLLTLGSSHGTGPGMTLQKLKSFNTNSYESSFLEKLCLLLHPSG	60
QY	1	MGSGPLVLLTLGSSHGTGPGMTLQKLKSFNTNSYESSFLEKLCLLLHPSG	60
Db	61	TSVTLHARSQHVVNCVT	78
QY	61	TSVTLHARSQHVVNCVT	78

```

RESULT      2
ID   W70327 standard; Protein; 46 AA.
AC   W70327;
DT   21-DEC-1998 (first entry)
DE   Secreted protein F878.1.
KW   Secreted protein; DU123.1; human.
OS   Homo sapiens.
FH   Key      Location/Qualifiers
FT   Peptide  8..20
FT   /note= "predicted leader/signal sequence, or
FT           transmembrane domain"
FT   PN       W09838209-A2.
FT   PD       03-SEP-1998.
FT   PR       25-FEB-1998: U03697.
FT   PR       24-FEB-1998: US-028724.
FT   PR       26-FEB-1997: US-805819.
FT   PA       (GENY ) GENETICS INST INC.
FT   PI       Agostino M, Jacobs K, Lavallee ER, McCoy JM, Merberg D,
FT   PI       Racie LA, Spaulding V, Treacy M;
FT   DR       N-PSDB; V33199.
FT   PT       New isolated polynucleotide(s) and encoded polypeptide(s) -
FT           obtained from human foetal kidney, adult colon, adult brain, foetal
FT           brain and placenta cDNA libraries.
FT   PS       Claim 36; Page 83; 103pp; English.
FT   CC       This is the amino acid sequence of novel human secreted protein
FT           F878.1, as deduced from a full-length cDNA clone (see V33199)
FT           obtained from a human adult placenta cDNA library. Database
FT           searching revealed some similarity between F878.1 and some known
FT           sequences. The invention provides new isolated polynucleotides
FT           (see V33199-99), from human foetal kidney, adult colon, adult brain,
FT           foetal brain and placenta cDNA libraries, that code for secreted
FT           proteins (see W70319-27). The clones can be used for recombinant
FT           production of the polypeptides, which may have activities such as
FT           e.g. nutritional activity, cytokine and cell proliferation or
FT           differentiation activity, immunostimulant or immunosuppressive,
FT           haematopoiesis regulating activity, tissue growth activity, activin
FT           or inhibin activity, chemotactic or chemokinetic activity,
FT           haemostatic and thrombolytic activity, receptor/ligand activity,
FT           antiinflammatory activity, cadherin/tumour invasion suppressor
FT           activity, tumour inhibition activity, or other activities.
FT   SQ       Sequence 46 AA;

Query Match      15.4%; Score 86; DB 1; Length 46;
Best Local Similarity 56.5%; Pred. No. 4.56e+01;
Matches 13; Conservative 6; Mismatches 3; Indels 1; Gaps 1;

Db   6 GAALPILLLLALRGTFHGARGP 28
QY   2 GSGLP-LVLLTLGSSHGTPG 23
      |::||| |::||| |::||| |::|||

RESULT      3
ID   W73408 standard; Protein; 47 AA.
AC   W73408;
DT   19-FEB-1999 (first entry)
DE   Human secreted protein encoded by Gene No. 12.
KW   Secreted protein; human; protein therapy; gene therapy; blood disorder;
KW   pathological condition; diagnosis; cancer; neurological disorder;
KW   developmental abnormality; foetal deficiency; leukaemia; hepatic disease;
KW   immune system disorder; Alzheimer's disease; cognitive disorder;
KW   schizophrenia; prostate disease; autoimmune disorder; AIDS.
OS   Homo sapiens.
FH   Key      Location/Qualifiers
FT   Misc_difference 47
FT   /note= "unspecified amino acid"
FT   PN       W09854206-A1.
FT   PD       03-DEC-1998.
FT   PR       28-MAY-1998: U10868.
FT   PR       29-AUG-1997: US-056296.
FT   PR       30-MAY-1997: US-044039.
FT   PR       30-MAY-1997: US-048093.

```

```

PR   30-MAY-1997; US-048101.
PR   30-MAY-1997; US-048190.
PR   30-MAY-1997; US-048356.
PR   30-MAY-1997; US-050935.
PR   29-AUG-1997; US-056250.
PR   29-AUG-1997; US-056293.
PR   (HUMA-) HUMAN GENOME SCI INC.
PI   Carter KC, Dillon PJ, Endress GA, Feng P, Ni J,
PI   Rosen CA, Ruben SM, Yu G;
PI   WPI; 99-070209/06.
DR   N-PSDB; Y08822.
DR   New isolated human genes - useful for diagnosis and treatment of,
PT   e.g. cancers, neurological disorders, immune diseases, developmental
PT   disorders or blood disorders
PT   Claim 11; Page 152-153; 188pp; English.
CC   This sequence is encoded by a cDNA of the invention, designated
CC   Gene No. 12. This sequence represents a human secreted protein, and is
CC   expressed in activated neutrophils, endothelial cells, T-cells and
CC   to a lesser extent in brain and liver.
CC   The DNA sequences of the invention and their corresponding secreted
CC   polypeptides are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Also pathological conditions
CC   can be diagnosed by determining the amount of the new polypeptides in a
CC   sample or by determining the presence of mutations in the DNA sequences.
CC   Specific uses are described for each of the DNA sequences and the encoded
CC   proteins, based on which tissues they are most highly expressed in, and
CC   include developing products for the diagnosis or treatment of cancer,
CC   tumours, neurological disorders, developmental abnormalities and foetal
CC   deficiencies, blood disorders, leukaemias, diseases of the immune system
CC   (including allergies or asthma), hepatic disease, Alzheimer's and
CC   cognitive disorders, schizophrenia, prostate disease, autoimmune
CC   disorders and AIDS. The polypeptides are also useful for identifying
CC   their binding partners.
SQ   Sequence 47 AA;

Query Match      15.4%; Score 86; DB 1; Length 47;
Best Local Similarity 56.5%; Pred. No. 4.56e+01;
Matches 13; Conservative 6; Mismatches 3; Indels 1; Gaps 1;

Db   6 GAALPILLLLALRGTFHGARGP 28
QY   2 GSGLP-LVLLTLGSSHGTPG 23
      |::||| |::||| |::||| |::|||

RESULT      4
ID   R48730 standard; Protein; 325 AA.
AC   R48730;
DT   06-JUN-1996 (first entry)
DE   G-protein coupled bovine adrenal angiotensin II type-1 receptor protein.
KW   G-protein coupled receptor; ligand binding assay; transmembrane domain;
KW   psychotic disorder; schizophrenia; dopamine; cAMP; adenosine; thrombin;
KW   muscarinic acetylcholine; adrenergic; endothelin; bombesin; endocrine;
KW   rhodopsin; opsin; odorant; cytomegalovirus.
OS   Bos taurus.
PN   W09405695-A1.
PD   17-MAR-1994.
PF   09-SEP-1993; U08528.
PR   10-SEP-1992; US-943236.
PR   (UNYK ) UNIV NEW YORK STATE.
PI   Murphy RB, Schuster DI;
PI   WPI; 94-101120/12.
DR   Polypeptides of G-coupled receptor proteins (GPRs) - useful for
PT   binding GPR ligands or modulating GPR binding
PT   Disclosure; Page 106-107; 160pp; English.
CC   Proteins R48685-R48758 represent a range of G-protein coupled receptor
CC   proteins selected from cAMP, adenosine, muscarinic acetylcholine, opsin,
CC   adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
CC   odorant, cytomegaloviral and other G-protein coupled receptors. The
CC   receptor proteins were used to design polypeptides, pref. based on the
CC   transmembrane domains, for use in G-protein coupled receptor ligand
CC   binding assays. The polypeptide fragments retain biological activity
CC   such as binding a GPR ligand or modulating GPR ligand binding to a GPR
CC   (see R48759-R48758, R50569-R50807 and R89189-R89195 for examples of

```


CC polypeptide fragments). The polypeptide fragments can be used in
 CC compositions for treating subjects suffering from a pathology related to
 CC a GPR abnormality e.g. a psychotic disorder such as schizophrenia.
 SQ Sequence 325 AA;

Query Match 15.2%; Score 85; DB 1; Length 325;
 Best Local Similarity 44.4%; Pred. No. 5.33e+01;
 Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLKYASVFLNLALADLCFLTLTP 53
 :: :| | | | | | | | | | | | | | | |

QY 34 FLTNSYESSFL-EL-LEKLCILLHLP 58
 :: :| | | | | | | | | | | | | | | |

RESULT 5

ID W02702 standard; peptide; 325 AA.

AC W02702;

DE 13-NOV-1996 (first entry)

DE G-protein coupled bovine adrenal angiotensin II type-1 receptor.
 KW G-protein coupled receptor; ligand binding assay; transmembrane domain;
 KW schizophrenia; dopamine; CAMP; adenosine; thrombin; adrenergic; opsin;
 KW muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;
 KW odorant; cytomegalovirus; serotonergic.
 OS Bos taurus.

PN US508384-A.

PD 16-APR-1996.

PF 10-SEP-1992; 943236.

PR 10-SEP-1992; US-943236.

PR 09-SEP-1993; US-118270.

PA (UYNV) UNIV NEW YORK STATE.

PI Murphy RB, Schuster DI;

DR WPI; 96-208785/21.

PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.
 PT for treating schizophrenia

PS Disclosure; Column 129-132; 184pp; English.

CC Proteins W02657-W02730 represent a range of G-protein coupled receptor
 CC (GPR) proteins selected from CAMP, adenosine, muscarinic acetylcholine,
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,
 CC odorant, cytomegaloviral and other GPR proteins. The receptor proteins
 CC were used to design polypeptides, pref. based on the transmembrane
 CC domains, for use in G-protein coupled receptor ligand binding assays.

CC The polypeptide fragments retain biological activity such as binding a
 CC GPR ligand or modulating GPR ligand binding to a GPR (see W02747-W02999
 CC for examples of polypeptide fragments). The polypeptide fragments can
 CC be used in compositions for treating subjects suffering from a pathology
 CC related to a GPR abnormality e.g. a psychotic disorder such as
 CC schizophrenia.

SQ Sequence 325 AA;

Query Match 15.2%; Score 85; DB 1; Length 325;
 Best Local Similarity 44.4%; Pred. No. 5.33e+01;
 Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLKYASVFLNLALADLCFLTLTP 53
 :: :| | | | | | | | | | | | | | | |

QY 34 FLTNSYESSFL-EL-LEKLCILLHLP 58
 :: :| | | | | | | | | | | | | | | |

RESULT 6

ID W98259 standard; Protein; 154 AA.

AC W98259;

DT 31-MAR-1999 (first entry)

DE H. pylori GHPO 895 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.

OS Helicobacter pylori.

PN W09843478-A1.

PD 08-OCT-1998.

PF 01-APR-1998; U06371.

PR 29-JUL-1997; US-902615.

PR 01-APR-1997; US-833457.

PR 24-JUN-1997; US-881227.

PA (HUMA-) HUMAN GENOME SCI INC.

PA (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI; 98-542293/46.
 DR N-PSDB; X13978.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 8; Page 250-251; 2054pp; English.
 CC This sequence represents a Helicobacter pylori GHPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SQ Sequence 154 AA;

Query Match 15.1%; Score 84; DB 1; Length 154;
 Best Local Similarity 46.9%; Pred. No. 6.24e+01;
 Matches 15; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Db 74 LGFGFGGLGLGLGLGFWFSFGSSPFSSFL 105
 || :| | | | | | | | | | | | | | | |

QY 14 LGSSTGTGPGMTLQLKLKESFLTNSYESSFL 45
 || :| | | | | | | | | | | | | | | |

RESULT 7

ID W97842 standard; Protein; 795 AA.

AC W97842;

DT 07-JUN-1999 (first entry)

DE Human P2Y11 receptor.

KW P2Y11; G protein coupled receptor; human; infection; neutropaenia;
 KW agranulocytosis; cancer; leukaemia; diagnosis; therapy.

OS Homo sapiens.

PH Key

FT Modified_site 474 Location/Qualifiers

FT /note= "putative protein kinase C phosphorylation
 FT site"

FT Modified_site 600 /note= "N-glycosylation"

FT Modified_site 794 /note= "calmodulin-dependent protein kinase
 FT phosphorylation site"

FT W09902675-A1.

PD 21-JAN-1999.

PF 09-JUL-1997; EP-870101.

PR 09-JUL-1997; BE0108.

PA (EURO-) EUROSREEN SA.

PI Roeynaems J, Communi D;

DR WPI; 99-120876/10.

DR N-PSDB; X07369.

PT New G protein-coupled receptor - useful for diagnosis, treatment and
 PT prevention of neutropaenia, agranulocytosis, infection and cancer
 PS Claim 2; Fig 1; 46pp; English.
 CC This polypeptide comprises a novel human G protein coupled receptor,
 CC termed p2Y11 that has selective affinity for ATP. The amino acid
 CC sequence was deduced from genomic DNA clones (see X07369). The
 CC invention also provides vectors, transformed cells, anti-p2Y11
 CC antibodies, nucleic acid probes, pharmaceutical compositions
 CC comprising such products and transgenic animals. Antisense
 CC nucleotides (claimed) that hybridise to mRNA are used to decrease
 CC activity of p2Y11, while specific antibodies are used to block
 CC binding of p2Y11 to its ligand. Probes are used in hybridisation
 CC assays to detect expression of p2Y11 at the RNA level, while
 CC antibodies are used similarly at the protein level in standard
 CC immunoassays, particularly for diagnosis of leukaemia. The
 CC transgenic animals are used to determine the effects of varying
 CC levels of p2Y11 expression. These animals, and host cells, are
 CC used in drug screening methods to identify (ant)agonists that are
 CC potentially useful for treatment or prevention of disorders
 CC associated with excessive or inadequate receptor activity.
 CC specifically neutropaenia, agranulocytosis, infections and cancer.
 CC Host cells are also used to produce recombinant p2Y11.


```

SQ      sequence 401 AA,
Query Match      14.5%; Score 81; DB 1; Length 401;
Best Local Similarity 38.1%; Pred. No. 9.94e+01;
Matches      8; Conservative      9; Mismatches      4; Indels      0; Caps      0;

Db      332 YFSSGFGQADRHLLRLICGL 352
      : : : : : : : | | | | | |
QY      34 FLINSSFFLEDEKICLT 54

RESULT      13
ID      W10038 standard: Protein: 1253 AA.

```

AC	W10036; 1998	(first entry)	
AD	07-FEB-1998		
DE	Mad binding protein, mSINa.		
KE	murine; mSINa; mammalian homologue; <i>Saccharomyces cerevisiae</i> ; repressor;		
DE	sin3; Mad; Max; mSIN:Mad complex; mSIN:Mad complex; MvC; promoter;		
KW	basic helix-loop-helix zipper protein; compete; DNA-binding;		
KW	Myc:Mad complex; activate; transcription; gene regulation.		
OS	Mus musculus.		

FT	Misc_difference 10	/label= unknown
FT		/note= "encoded by TAG"
FT	Misc_difference 1238	/label= unknown
FT		/note= "encoded by TAG"
PD	US5624818-A.	
PN	29-APR-1997.	
PD	01-JUN-1994; 252966.	
PF	01-JUN-1994; US-252966.	
PR	19-SEP-1991; US-756195.	
PR	23-JUN-1992; US-903710.	
PR	01-APR-1994; US-222638.	
PR	(HUTC-) HUTCHINSON CANCER RES CENT FRED.	
PI	AYER DE, Eisenman RN;	
PI	WPI: 97-258216/23.	
DR	N-PSDS; T70126.	
DR	mSln nucleic acids encoding recombinant polypeptide(s) that	
PT	associate with Mad polypeptide - are possible homologues of S.	
PT	cerevisiae general repressor protein Sin3	
PT	Example 12; F19 23A-C; il1bp; English.	
PS	This sequence represents the murine protein, designated mSlnA, which may	
CC	be a mammalian homologue of the Saccharomyces cerevisiae general	
CC	repressor protein Sin3. The mSln protein associates with a Mad	
CC	polypeptide to form a mSln:Mad complex, which preferably associates	
CC	with a Max polypeptide to form a mSln:Mad:Max complex which binds to a	
CC	nucleotide sequence comprising CACGTG. Mad is a basic helix-loop-helix	
CC	(bHLH) zipper protein which can compete with Myc by forming sequence-	
CC	specific DNA-binding heterocomplexes with Max. Mad:Max complexes repress	
CC	while Myc:Max complexes activate, transcription from promoters containing	
CC	proximal CACGTG binding sites for these proteins. Expression of Mad is	
CC	closely linked to differentiation in at least two distinct cell lineages.	
CC	The switch from Myc:Max to Mad:Max complexes may reflect the repression	
CC	of transcription of Myc regulated genes by Mad. The DNA, vectors and host	
CC	cells of the invention are useful for the recombinant production of mSln	
CC	proteins useful in elucidation of Mad repressor functions.	
SO	Sequence 1253 AA:	

Accession	Protein Name	Location/Qualifiers	Query Match	Score	DB 1	Length	DB 2	Score	DB 3	Length	DB 4	Score	DB 5	Length	DB 6	Score	DB 7	Length	DB 8	Score	DB 9	Length	DB 10	Score	DB 11	Length	DB 12	Score	DB 13	Length	DB 14	Score	DB 15	Length	DB 16	Score	DB 17	Length	DB 18	Score	DB 19	Length	DB 20	Score	DB 21	Length	DB 22	Score	DB 23	Length	DB 24	Score	DB 25	Length	DB 26	Score	DB 27	Length	DB 28	Score	DB 29	Length	DB 30	Score	DB 31	Length	DB 32	Score	DB 33	Length	DB 34	Score	DB 35	Length	DB 36	Score	DB 37	Length	DB 38	Score	DB 39	Length	DB 40	Score	DB 41	Length	DB 42	Score	DB 43	Length	DB 44	Score	DB 45	Length	DB 46	Score	DB 47	Length	DB 48	Score	DB 49	Length	DB 50	Score	DB 51	Length	DB 52	Score	DB 53	Length	DB 54	Score	DB 55	Length	DB 56	Score	DB 57	Length	DB 58	Score	DB 59	Length	DB 60	Score	DB 61	Length	DB 62	Score	DB 63	Length	DB 64	Score	DB 65	Length	DB 66	Score	DB 67	Length	DB 68	Score	DB 69	Length	DB 70	Score	DB 71	Length	DB 72	Score	DB 73	Length	DB 74	Score	DB 75	Length	DB 76	Score	DB 77	Length	DB 78	Score	DB 79	Length	DB 80	Score	DB 81	Length	DB 82	Score	DB 83	Length	DB 84	Score	DB 85	Length	DB 86	Score	DB 87	Length	DB 88	Score	DB 89	Length	DB 90	Score	DB 91	Length	DB 92	Score	DB 93	Length	DB 94	Score	DB 95	Length	DB 96	Score	DB 97	Length	DB 98	Score	DB 99	Length	DB 100	Score	DB 101	Length	DB 102	Score	DB 103	Length	DB 104	Score	DB 105	Length	DB 106	Score	DB 107	Length	DB 108	Score	DB 109	Length	DB 110	Score	DB 111	Length	DB 112	Score	DB 113	Length	DB 114	Score	DB 115	Length	DB 116	Score	DB 117	Length	DB 118	Score	DB 119	Length	DB 120	Score	DB 121	Length	DB 122	Score	DB 123	Length	DB 124	Score	DB 125	Length	DB 126	Score	DB 127	Length	DB 128	Score	DB 129	Length	DB 130	Score	DB 131	Length	DB 132	Score	DB 133	Length	DB 134	Score	DB 135	Length	DB 136	Score	DB 137	Length	DB 138	Score	DB 139	Length	DB 140	Score	DB 141	Length	DB 142	Score	DB 143	Length	DB 144	Score	DB 145	Length	DB 146	Score	DB 147	Length	DB 148	Score	DB 149	Length	DB 150	Score	DB 151	Length	DB 152	Score	DB 153	Length	DB 154	Score	DB 155	Length	DB 156	Score	DB 157	Length	DB 158	Score	DB 159	Length	DB 160	Score	DB 161	Length	DB 162	Score	DB 163	Length	DB 164	Score	DB 165	Length	DB 166	Score	DB 167	Length	DB 168	Score	DB 169	Length	DB 170	Score	DB 171	Length	DB 172	Score	DB 173	Length	DB 174	Score	DB 175	Length	DB 176	Score	DB 177	Length	DB 178	Score	DB 179	Length	DB 180	Score	DB 181	Length	DB 182	Score	DB 183	Length	DB 184	Score	DB 185	Length	DB 186	Score	DB 187	Length	DB 188	Score	DB 189	Length	DB 190	Score	DB 191	Length	DB 192	Score	DB 193	Length	DB 194	Score	DB 195	Length	DB 196	Score	DB 197	Length	DB 198	Score	DB 199	Length	DB 200	Score	DB 201	Length	DB 202	Score	DB 203	Length	DB 204	Score	DB 205	Length	DB 206	Score	DB 207	Length	DB 208	Score	DB 209	Length	DB 210	Score	DB 211	Length	DB 212	Score	DB 213	Length	DB 214	Score	DB 215	Length	DB 216	Score	DB 217	Length	DB 218	Score	DB 219	Length	DB 220	Score	DB 221	Length	DB 222	Score	DB 223	Length	DB 224	Score	DB 225	Length	DB 226	Score	DB 227	Length	DB 228	Score	DB 229	Length	DB 230	Score	DB 231
-----------	--------------	---------------------	-------------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	------	-------	------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	-------	-------	-------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------	--------	--------	-------	--------

WORLD

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 20 12:33:19 2000; MasPar time 4.10 Seconds
Tabular output not generated. 246.359 Million cell updates/sec

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSLPLVLLTLGSSHG.....SGTSTLHARSHHVVCNT 78

Scoring table:

PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 25.603; Variance 123.523; scale 0.207

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	85	15.2	325	1	US-08-118-Sequence 51, Applicati	2.28e+01
2	85	15.2	325	3	PCT-US93-0Sequence 51, Applicati	2.28e+01
3	84	15.1	374	2	US-09-009-Sequence 2, Applicatio	2.68e+01
4	83	14.9	194	1	US-08-063-Sequence 8, Applicatio	3.14e+01
5	83	14.9	194	3	PCT-US93-0Sequence 8, Applicatio	3.14e+01
6	82	14.7	169	1	US-08-741-Sequence 2, Applicatio	3.68e+01
7	81	14.5	401	2	US-08-820-Sequence 2, Applicatio	4.31e+01
8	81	14.5	1253	1	US-08-252-Sequence 12, Applicati	4.31e+01
9	81	14.5	1261	1	US-08-252-Sequence 18, Applicati	4.31e+01
10	80	14.3	169	2	US-08-828-Sequence 4, Applicatio	5.04e+01
11	80	14.3	390	1	US-08-117-Sequence 6, Applicatio	5.04e+01
12	80	14.3	390	3	PCT-US93-0Sequence 6, Applicatio	5.04e+01
13	80	14.3	390	1	US-08-216-Sequence 6, Applicatio	5.04e+01
14	80	14.3	390	1	US-07-817-Sequence 6, Applicatio	5.04e+01
15	80	14.3	398	1	US-08-542-Sequence 6, Applicatio	5.04e+01
16	80	14.3	398	1	US-08-370-Sequence 6, Applicatio	5.04e+01
17	80	14.3	1214	2	US-08-486-Sequence 54, Applicati	5.04e+01
18	80	14.3	1214	2	US-08-231-Sequence 54, Applicati	5.04e+01
19	80	14.3	1219	2	US-08-231-Sequence 50, Applicati	5.04e+01
20	80	14.3	1219	2	US-08-486-Sequence 50, Applicati	5.04e+01
21	80	14.3	1231	2	US-08-231-Sequence 48, Applicati	5.04e+01
22	80	14.3	1231	2	US-08-486-Sequence 48, Applicati	5.04e+01
23	80	14.3	1236	2	US-08-486-Sequence 6, Applicatio	5.04e+01

24	80	14.3	1236	2	US-08-231-Sequence 6, Applicatio	5.04e+01
25	80	14.3	1239	2	US-08-486-Sequence 52, Applicati	5.04e+01
26	80	14.3	1239	2	US-08-231-Sequence 52, Applicati	5.04e+01
27	80	14.3	1244	2	US-08-486-Sequence 46, Applicati	5.04e+01
28	80	14.3	1244	2	US-08-231-Sequence 46, Applicati	5.04e+01
29	79	14.2	429	1	US-08-339-Sequence 33, Applicati	5.90e+01
30	79	14.2	652	2	US-08-751-Sequence 2, Applicatio	5.90e+01
31	79	14.2	969	2	US-08-548-Sequence 1, Applicatio	5.90e+01
32	79	14.2	986	2	US-08-548-Sequence 3, Applicatio	5.90e+01
33	78	14.0	120	1	US-08-420-Sequence 29, Applicati	6.90e+01
34	78	14.0	120	3	PCT-US95-1Sequence 29, Applicati	6.90e+01
35	78	14.0	1239	1	US-08-026-Sequence 3, Applicatio	6.90e+01
36	76	13.6	197	2	US-08-215-Sequence 3, Applicatio	9.41e+01
37	76	13.6	317	2	US-09-213-Sequence 3, Applicatio	9.41e+01
38	76	13.6	340	2	US-09-213-Sequence 1, Applicatio	9.41e+01
39	76	13.6	477	4	5245013-3Patent No. 5245013	9.41e+01
40	76	13.6	477	2	US-08-990-Sequence 2, Applicatio	9.41e+01
41	76	13.6	481	2	US-08-779-Sequence 2, Applicatio	9.41e+01
42	76	13.6	481	1	US-08-261-Sequence 4, Applicatio	9.41e+01
43	76	13.6	481	1	US-08-186-Sequence 2, Applicatio	9.41e+01
44	76	13.6	481	2	US-08-215-Sequence 9, Applicatio	9.41e+01
45	76	13.6	481	1	US-08-306-Sequence 98, Applicati	9.41e+01

ALIGNMENTS

RESULT 1	US-08-118-270-51	STANDARD;	PRT;	325 AA.
ID XX	xxxxxx			
AC XX				
DT XX				
DE XX	Sequence 51, Application US/08118270			
CC CC	Sequence 51, Application US/08118270			
CC CC	Patent No. 5508384			
CC CC	GENERAL INFORMATION:			
CC CC	APPLICANT: Murphy, Randall B.			
CC CC	APPLICANT: Schuster, David I.			
CC CC	TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN			
CC CC	TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF			
CC CC	NUMBER OF SEQUENCES: 348			
CC CC	CORRESPONDENCE ADDRESS:			
CC CC	ADDRESSEE: BROWDY AND NEIMARK			
CC CC	STREET: 419 Seventh Street, N.W., Suite 300			
CC CC	CITY: Washington			
CC CC	STATE: D.C.			
CC CC	COUNTRY: USA			
CC CC	ZIP: 20004			
CC CC	COMPUTER READABLE FORM:			
CC CC	MEDIUM TYPE: Floppy disk			
CC CC	COMPUTER: IBM PC compatible			
CC CC	OPERATING SYSTEM: PC-DOS/MS-DOS			
CC CC	SOFTWARE: PatentIn Release #1.0, Version #1.25			
CC CC	CURRENT APPLICATION DATA:			
CC CC	APPLICATION NUMBER: US/08/118,270			
CC CC	FILING DATE: 09-SEP-1993			
CC CC	PRIOR APPLICATION DATA:			
CC CC	APPLICATION NUMBER: US 07/943,236			
CC CC	FILING DATE: 10-SEP-1992			
CC CC	ATTORNEY/AGENT INFORMATION:			
CC CC	NAME: Townsend, Kevin G.			
CC CC	REGISTRATION NUMBER: 34,033			
CC CC	REFERENCE/DOCKET NUMBER: MURPHY-2A			
CC CC	TELECOMMUNICATION INFORMATION:			
CC CC	TELEPHONE: 202-628-5197			
CC CC	TELEFAX: 202-737-3528			
CC CC	TELEX: 248633			
CC CC	INFORMATION FOR SEQ ID NO: 51:			
CC CC	SEQUENCE CHARACTERISTICS:			
CC CC	LENGTH: 325 amino acids			
CC CC	TYPE: amino acid			

```
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 325 AA; 37288 MW; 592694 CN;

Query Match
Best Local Similarity 15.2%; Score 85; DB 1; Length 325;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLTYSVFLNLAADLCFLITLP 53
:: :| | | :| | | | | | | |
QY 34 FLTNSYESSFL-EL-LKLCILLHLP 58

RESULT 2
ID PCT-US93-08528-51 STANDARD; PRT; 325 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 51, Application PC/TUS9308528
CC CC
CC GENERAL INFORMATION:
CC APPLICANT: New York University
CC TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
CC NUMBER OF SEQUENCES: 348
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: BROWDY AND NEIMARK
CC STREET: 419 Seventh Street, N.W., Suite 300
CC CITY: Washington
CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20004
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION NUMBER: PCT/US93/08528
CC FILING DATE: 09-SEP-1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 07/943,236
CC FILING DATE: 10-SEP-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Townsend, Kevin G.
CC REGISTRATION NUMBER: 34,033
CC REFERENCE/DOCKET NUMBER: MURPHY-2 PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-628-5197
CC TELEFAX: 202-737-3528
CC TELEX: 248633
CC INFORMATION FOR SEQ ID NO: 51:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 325 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 325 AA; 37288 MW; 592694 CN;

Query Match
Best Local Similarity 15.2%; Score 85; DB 3; Length 325;
Matches 12; Conservative 5; Mismatches 8; Indels 2; Gaps 2;

Db 27 YMKLTYSVFLNLAADLCFLITLP 53
:: :| | | :| | | | | | | |
QY 34 FLTNSYESSFL-EL-LKLCILLHLP 58
```

```
RESULT 3
ID US-09-009-438-2 STANDARD; PRT; 374 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 2, Application US/09009438
CC CC
CC SEQUENCE 2, Application US/09009438
CC Patent No. 5981223
CC GENERAL INFORMATION:
CC APPLICANT: SATHE, GANESH M.
CC APPLICANT: HALSEY, WENDY S.
CC APPLICANT: BERGSMAN, DEREK J.
CC TITLE OF INVENTION: HDXU17 ENCODES A NOVEL 7-TRANSMEMBRANE
CC NUMBER OF SEQUENCES: 2
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: RATNER & PRESTIA
CC STREET: P.O. BOX 980
CC CITY: VALLEY FORGE
CC STATE: PA
CC COUNTRY: USA
CC ZIP: 19482
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/09/009,438
CC FILING DATE: 20-JAN-1998
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: PRESTIA, PAUL F
CC REGISTRATION NUMBER: 23,031
CC REFERENCE/DOCKET NUMBER: GH-70369
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 610-407-0700
CC TELEFAX: 610-407-0701
CC TELEX: 846169
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 374 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 374 AA; 40345 MW; 690567 CN;

Query Match
Best Local Similarity 15.1%; Score 84; DB 2; Length 374;
Matches 16; Conservative 12; Mismatches 11; Indels 3; Gaps 2;

Db 212 LGGCLPLLLTLTAAYGALGRAVLSPGTMVAEKLRAALVAGS 253
:| | | | | :| | | | | :| | | | | :| | | | |
QY 1 MGSGLPLVLLLTLLGS-SHGT--GPGMTLQLKAKESFLTNS 39

RESULT 4
ID US-08-063-552-8 STANDARD; PRT; 194 AA.
XX AC xxxxxx
XX DT
XX DE
XX SEQUENCE 8, Application US/08063552
CC CC
CC SEQUENCE 8, Application US/08063552
```

CC Patent No. 5688936
CC GENERAL INFORMATION:
CC APPLICANT: Edwards, Robert H
CC TITLE OF INVENTION: Vesicle Membrane Transport Proteins
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Sheldon & Mak
CC STREET: 225 South Lake Avenue, Ninth Floor
CC CITY: Pasadena
CC STATE: California
CC COUNTRY: USA
CC ZIP: 91101
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC FILING DATE: US/08/063,552
CC APPLICATION NUMBER: 19930514
CC CLASSIFICATION: 530
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Farber, Michael B
CC REGISTRATION NUMBER: 32,612
CC REFERENCE/DOCKET NUMBER: 9067-1
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (818) 796-4000
CC TELEFAX: (818) 795-6321
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 194 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: Transposon 10
CC SEQUENCE 194 AA; 20868 MW; 201442 CN;

Query Match 14.9%; Score 83; DB 1; Length 194;
Best Local Similarity 31.0%; Pred. No. 3.14e+01;
Matches 18; Conservative 13; Mismatches 23; Indels 4; Gaps 4;
Db 1 MNSSTKIALVTLLDAM-GIGLIMPVLTLLRE-FIASEDIANHFVLLA-LYALMQV 55
QY 1 MGSGLPLVLLLTLLGSSHGTPGWT-LQLKLESFLTNSYESSFLELLEKLCLLHL 57

RESULT 5
ID PCT-US93-05704-8 STANDARD; PRT; 194 AA.
XX
XX
AC xxxxxx
XX
DT
XX
XX
DE Sequence 8, Application PC/TUS9305704
CC
CC GENERAL INFORMATION:
CC APPLICANT: Edwards, Robert H
CC TITLE OF INVENTION: Vesicle Membrane Transport Proteins
CC NUMBER OF SEQUENCES: 17
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Sheldon & Mak
CC STREET: 225 South Lake Avenue, Ninth Floor
CC CITY: Pasadena
CC STATE: California
CC COUNTRY: USA
CC ZIP: 91001
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC . COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/05704
CC FILING DATE: 19930611
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Farber, Michael B
CC REGISTRATION NUMBER: 32,612
CC REFERENCE/DOCKET NUMBER: 9067-1PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (818) 796-4000
CC TELEFAX: (818) 795-6321
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 194 amino acids
CC TYPE: AMINO ACID
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
CC ORIGINAL SOURCE:
CC ORGANISM: Transposon 10
CC SEQUENCE 194 AA; 20868 MW; 201442 CN;
SQ
Query Match 14.9%; Score 83; DB 3; Length 194;
Best Local Similarity 31.0%; Pred. No. 3.14e+01;
Matches 18; Conservative 13; Mismatches 23; Indels 4; Gaps 4;
Db 1 MNSSTKIALVTLLDAM-GIGLIMPVLTLLRE-FIASEDIANHFVLLA-LYALMQV 55
QY 1 MGSGLPLVLLLTLLGSSHGTPGWT-LQLKLESFLTNSYESSFLELLEKLCLLHL 57
RESULT 6
ID US-08-741-406-2 STANDARD; PRT; 169 AA.
XX
XX
AC xxxxxx
XX
DT
XX
DE Sequence 2, Application US/08741406
XX
CC Sequence 2, Application US/08741406
CC Patent No. 5721118
CC GENERAL INFORMATION:
CC APPLICANT: Scheffler, Immo E.
CC TITLE OF INVENTION: Mammalian Artificial Chromosomes and
CC TITLE OF INVENTION: Methods of Using Same
CC NUMBER OF SEQUENCES: 16
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell & Flores LLP
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: United States
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/741,406
CC FILING DATE:
CC CLASSIFICATION: 514
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/550,717
CC FILING DATE: 31-OCT-1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-UD 2317

CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WordPerfect, Version 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/117,006
CC FILING DATE:
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;
SQ
Query Match 14.3%; Score 80; DB 1; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
Db 52 LVMLALITLTLNAPVIATVTRKLTLPANYLIASLDVTLVSLVPISTMTV 111
QY 7 LVLLTLGSSHGTPGWTQLKKEFL-TNSSVESFLELEKLCILLHLPSGTSVTL 65
Db 112 TDRWTLQVVD 123
QY 66 HHARSQHVVCN 77
RESULT 12
ID PCT-US93-00149-6 STANDARD; PRT; 390 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 6, Application PC/TUS9300149
XX
Sequence 6, Application PC/TUS9300149
CC GENERAL INFORMATION:
CC APPLICANT: Weinshank, Richard L.
CC APPLICANT: Branchek, Theresa
CC TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/00149
CC FILING DATE: 19930108
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;
SQ
Query Match 14.3%; Score 80; DB 3; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
Db 52 LVMLALITLTLNAPVIATVTRKLTLPANYLIASLDVTLVSLVPISTMTV 111
QY 7 LVLLTLGSSHGTPGWTQLKKEFL-TNSSVESFLELEKLCILLHLPSGTSVTL 65
Db 112 TDRWTLQVVD 123
QY 66 HHARSQHVVCN 77
RESULT 13
ID US-08-216-594-6 STANDARD; PRT; 390 AA.
XX
AC xxxxxx
XX
DT
DE
Sequence 6, Application US/08216594
XX
Sequence 6, Application US/08216594
CC Patent No. 5652113
CC GENERAL INFORMATION:
CC APPLICANT: Weinshank, Richard L.
CC APPLICANT: Branchek, Theresa
CC TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/216,594
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;

Query Match 14.3%; Score 80; DB 1; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;
Db 52 LVMLALITLTIATLSNAFVIATYTRKLTTPANYLIASLDVTDLLVSLVPISTMTV 111
QY 7 LVLLTLGSSHGFGMTQLKLESFL-TNSSTESSFLELLEKLCLLHLPSTSVTL 65
Db 112 TDRWLSQVVD 123
QY 66 HHARSQHHVVCN 77

RESULT 14
ID US-07-817-920-6 STANDARD; PRT; 390 AA.

XX XXXXXX

XX XXXXXX

XX XXXXXX

Sequence 6, Application US/07817920

Sequence 6, Application US/07817920

Patent No. 5360735

GENERAL INFORMATION:

APPLICANT: Weinshank, Richard L

APPLICANT: Branchek, Theresa

APPLICANT: Hartig, Paul R

TITLE OF INVENTION: DNA ENCODING A HUMAN 5-HT1F RECEPTOR AND
USES THEREOF

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/817,920

CC FILING DATE: 19920108
CC CLASSIFICATION: 514
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/39318
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-977-9550
CC TELEFAX: 212-664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 390 amino acids
CC TYPE: AMINO ACID
CC STRANDEDNESS: unknown
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC IMMEDIATE SOURCE:
CC CLONE: 5-HT1DB
CC SEQUENCE 390 AA; 43656 MW; 849817 CN;

Query Match 14.3%; Score 80; DB 1; Length 390;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;

Db 52 LVMLALITLTIATLSNAFVIATYTRKLTTPANYLIASLDVTDLLVSLVPISTMTV 111
QY 7 LVLLTLGSSHGFGMTQLKLESFL-TNSSTESSFLELLEKLCLLHLPSTSVTL 65

Db 112 TDRWLSQVVD 123

QY 66 HHARSQHHVVCN 77

RESULT 15
ID US-08-542-358-6 STANDARD; PRT; 398 AA.

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

XX XXXXXX

CC REFERENCE/DOCKET NUMBER: 1795/39317-22/JPW/NAT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 212-278-0400
CC TELEFAX: 212-391-0525
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 398 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC MOLECULE TYPE: protein
CC HYPOTHEetical: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC SEQUENCE 398 AA; 44384 MW; 880684 CN;
SQ

Query Match 14.3%; Score 80; DB 1; Length 398;
Best Local Similarity 23.6%; Pred. No. 5.04e+01;
Matches 17; Conservative 22; Mismatches 32; Indels 1; Gaps 1;

Db 52 LYMLALITLTLTSLNAFYIATVIRKLTHTPANTYLIASLDVTLVSIPISTMTYV 111
QY 7 LVLLITLIGSSHGTPGWTQLKLESFL-TNSSYESSFLELEKLCLLHLPSTSVTL 65

Db 112 TDRWTLQVYCD 123
QY 66 HHARSQHVVVN 77

Search completed: Mon Mar 20 12:33:28 2000
Job time : 9 secs.

[M][P][E][R][C][H]

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPPerch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 20 12:32:21 2000; MasPar time 8.96 Seconds
Tabular output not generated. 410.868 Million cell updates/sec

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pap
Perfect Score: 558
Sequence: 1 MGSGLPLVLLLLGSSHGT.....SGTSVTLFHARSHQHVVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: p1r62
1:p1r1 2:p1r2 3:p1r3 4:p1r4

Statistics: Mean 39.408; Variance 84.670; scale 0.465

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length	ID	Description
1	101	18.1	501	1	C70384
2	93	16.7	382	2	S63638
3	92	16.5	712	2	A48156
4	91	16.3	252	2	JC5237
5	89	15.9	879	2	S73757
6	89	15.9	3005	2	S33642
7	88	15.8	314	2	B75076
8	88	15.8	1822	2	S44849
9	87	15.6	143	2	S43071
10	87	15.6	396	2	S58161
11	86	15.4	272	2	S39641
12	85	15.2	315	2	H64082
13	84	15.1	441	2	S74800
14	84	15.1	454	2	T03130
15	84	15.1	509	2	S76731
16	83	14.9	378	2	D70324
17	83	14.9	366	2	T14243
18	83	14.9	401	1	YR2C10
19	83	14.9	469	2	G65058
20	83	14.9	647	2	A37086
21	82	14.7	56	2	S53000
22	82	14.7	110	2	S22896
23	82	14.7	110	2	S23368

24	82	14.7	152	2	S58125	T-cell antigen recept	1.89e+01
25	82	14.7	155	2	S58178	T cell antigen recept	1.89e+01
26	82	14.7	372	2	A46138	invasion protein inve	1.89e+01
27	82	14.7	612	2	T14101	NADH dehydrogenase su	1.89e+01
28	82	14.7	946	2	T00024	ent-Kaurene synthase	1.89e+01
29	82	14.7	5825	2	T12117	polyprotein - fava be	1.89e+01
30	81	14.5	253	2	S70170	nota protein - Rhodob	2.46e+01
31	81	14.5	261	2	I52518	sperm acrosome antige	2.46e+01
32	81	14.5	287	2	S71192	mitosis-specific cycl	2.46e+01
33	81	14.5	346	2	JC5715	G protein-coupled rec	2.46e+01
34	81	14.5	346	2	JC5716	G protein-coupled rec	2.46e+01
35	81	14.5	369	2	D72238	peptide chain release	2.46e+01
36	81	14.5	387	1	A48326	ubiquinol--cytochrome	2.46e+01
37	81	14.5	469	2	H70626	probable narK3 protei	2.46e+01
38	81	14.5	474	2	T00943	hypothetical protein	2.46e+01
39	81	14.5	482	2	B35843	lipopolysaccharide-bi	2.46e+01
40	81	14.5	1215	2	B72029	swi/snf family helica	2.46e+01
41	81	14.5	1219	2	I61713	co-repressor protein	2.46e+01
42	81	14.5	1229	2	A56068	co-repressor protein	2.46e+01
43	80	14.3	191	2	H71370	hypothetical protein	3.19e+01
44	80	14.3	544	2	T16374	hypothetical protein	3.19e+01
45	80	14.3	741	2	I48694	probable transcriptio	3.19e+01

ALIGNMENTS

RESULT 1

ENTRY C70384 #type complete

TITLE protein export membrane protein SecD - Aquifex aeolicus

ORGANISM #formal_name Aquifex aeolicus

DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS C70384

REFERENCE A70300

#authors Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Sneed, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.

#journal Nature (1998) 392:353-358

#title The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

#cross-references MUID:98196666

#accession C70384

#status preliminary; nucleic acid sequence not shown; translation not shown

#molecule_type DNA

#residues 1-501 #label AQF

#cross-references GB:AE000716; MID:g2983478; PID:g2983481; GB:AE000657

#experimental_source strain VF5

GENETICS

#gene secD

CLASSIFICATION #superfamily protein export membrane protein secD

SUMMARY #length 501 #molecular-weight 55459 #checksum 2250

Query Match 18.1%; Score 101; DB 1; Length 501;

Best Local Similarity 38.3%; Pred. No. 8.92e-02;

Matches 18; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

Db 454 VILFQF-GSGPVKGFATLALGFIASFISVNYAKVFLDLNLSKIL 499

QY 8 VLLTLTGSSHGTPGWTQLKLKESFLTNSSYSFLELLEKICLL 54

RESULT 2

ENTRY S63638 #type complete

TITLE ubiquinol--cytochrome-c reductase (EC 1.10.2.2) cytochrome b

ALTERNATE_NAMES apocytochrome b

ORGANISM #formal_name mitochondrion Allomyces macrogynus

DATE 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 20-Sep-1999

ACCESSIONS S63638

REFERENCE S63635

```

#authors Paquin, B.; Lang, B.F.
#journal J. Mol. Biol. (1996) 255:688-701
#title The mitochondrial DNA of Allomyces macrognus: the complete genomic sequence from an ancestral fungus.
#cross-references EMBL:96226032
#accession S63638
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-362 #label PAQ
##cross-references EMBL:041288; NID:g1236403; PIDN:AAC49221.1;
PID:g1236404
##note the nucleotide sequence was submitted to the EMBL Data Library, November 1995

GENETICS
#gene cob
#gene mitochondrion
#introns 67/3; 137/3; 143/3; 164/1; 200/2; 252/3
CLASSIFICATION
#superfamily cytochrome b; cytochrome b homology; cytochrome b6 homology; plastocyanin--plastocyanin reductase 17K protein homology
KEYWORDS
electron transfer; heme; iron; mitochondrion; oxidative phosphorylation; oxidoreductase; respiratory chain
FEATURE
10-340 #domain cytochrome b homology #label CBH\
10-210 #domain cytochrome b6 homology #label CB6\
222-340 #domain plastocyanin--plastocyanin reductase 17K protein homology #label 17K\
82,197 #binding_site heme iron, low potential (His) (axial ligands) #status predicted\
96,202 #binding_site heme iron, high potential (His) (axial ligands) #status predicted\
SUMMARY
#length 382 #molecular-weight 43467 #checksum 6973
Query Match 16.7%; Score 93; DB 2; Length 382;
Best Local Similarity 33.3%; Pred. No. 9,25e-01;
Matches 17; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Db 8 PVLISANSLDPSFNITLWNFGSLG-LCLVIGIVGTGLAMHYAPS 57
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 22 PQMTLQK-LKESFL-TNSSEVFLELEKICLLHLPFGSVTLHHARS 70

RESULT 3
ENTRY A48156 #type complete
TITLE translation regulator GCD6 - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES guanine nucleotide exchange factor chain GCD6; protein YD8142B.03; protein YDR211w; translation initiation factor eIF-2B homology
DATE #formal_name Saccharomyces cerevisiae
ORGANISM 28-May-1993 sequence_revision 03-May-1994 #text_change
06-Feb-1998
ACCESSIONS A48156; S61578; S30776
REFERENCE A48156
#authors Bushman, J.L.; Asuru, A.I.; Matts, R.L.; Hinnebusch, A.G.
#journal Mol. Cell. Biol. (1993) 13:1920-1932
#title Evidence that GCD6 and GCD7, translational regulators of GCNA, are subunits of the guanine nucleotide exchange factor for eIF-2 in Saccharomyces cerevisiae.
#cross-references NID:93180841
#accession A48156
##molecule_type DNA
##residues 1-712 #label BUS
##cross-references EMBL:L07115; NID:g171572; PID:g171574
##note sequence extracted from NCBI backbone (NCBIN:126018, NCBI:P126021)

REFERENCE S61578
#authors Oliver, K.; Harris, D.
#submission submitted to the EMBL Data Library, December 1995
#accession S61578
##molecule_type DNA
##residues 1-712 #label OLI
##cross-references EMBL:P269195; NID:g1122341; PID:e213795; PID:g1122344; NID:YDR211w

```



```

##status      nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues    1-143 #label THO
##cross-references EMBL:X73675; NID:g469952; PID:g469957
##note        the nucleotide sequence was submitted to the EMBL Data
              Library, June 1993
SUMMARY       #length 143 #molecular-weight 13317 #checksum 6533

Query Match   15.6%; Score 87; DB 2; Length 143;
Best Local Similarity 47.2%; Pred. No. 4.95e+00;
Matches 17; Conservative 4; Mismatches 14; Indels 1; Gaps 1;

Db 88 LGUGLGLGGLGGLGGLGGLGGLGGLGGLGGLGGLGGL 123
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1 MGSGLPLVLLTL-LGSSHGTPGWTQLKLKESFL 35

RESULT 10
ENTRY   S58161      #type complete
TITLE   Probable translation releasing factor RF-1 - fission yeast
ALTERNATE_NAMES SPAC3F7.17 protein
ORGANISM Schizosaccharomyces pombe
DATE     13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change
        26-Aug-1999
ACCESSIONS S58161
REFERENCE  S58145
#authors   Gentles, S.; Churcher, C.M.
#submission submitted to the EMBL Data Library, July 1995
#accession S58161

##status      preliminary
##molecule_type DNA
##residues    1-396 #label GEN
##cross-references EMBL:Z50142; NID:g1052783; PIDN:CAA90504.1;
              PID:g1052800

GENETICS
#introns     57/3; 90/3
CLASSIFICATION #superfamily translation releasing factor
SUMMARY       #length 396 #molecular-weight 44954 #checksum 9362

Query Match   15.6%; Score 87; DB 2; Length 396;
Best Local Similarity 45.0%; Pred. No. 4.95e+00;
Matches 9; Conservative 8; Mismatches 3; Indels 0; Gaps 0;

Db 281 LTHIPTGTVSQDSRSQHQ 300
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 54 LLHLPSTGIVLHARSQHH 73

RESULT 11
ENTRY   S39641      #type complete
TITLE   flagellar motor apparatus homolog ytxD - Bacillus subtilis
ORGANISM #formal name Bacillus subtilis
DATE     08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change
        23-Sep-1999
ACCESSIONS S39641; B70003
REFERENCE  S39641
#authors   Grundy, F.J.; Waters, D.A.; Takova, T.Y.; Henkin, T.M.
#journal   Mol. Microbiol. (1993) 10:259-271
#title     Identification of genes involved in utilization of acetate
              and acetoin in Bacillus subtilis.
#cross-references MUID:95020526
#accession S39641

##status      preliminary
##molecule_type DNA
##residues    1-272 #label GRU
REFERENCE  A69580
#authors   Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;
              Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.;
              Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans,
              A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.;
              Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
              Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.;

```

```

Daniel, R.A.; Denizot, F.; Devine, K.M.; Duesterhoeft, A.;
Ehrlich, S.D.; Emmerison, P.T.; Entian, K.D.; Errington, J.;
Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita,
M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim,
S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.;
Guiseppi, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood,
C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.;
Hulio, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.;
Kashahara, Y.; Klaer-Blanchard, M.; Klein, C.; Kobayashi,
Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.;
Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.;
Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Mauel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno,
M.; Moesti, D.; Nakai, S.; Noback, M.; Noone, D.; O'Neill,
M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro,
V.; Pohl, T.M.; Portetelle, D.; Porwolik, S.; Prescott,
A.M.; Presecan, E.; Fujic, P.; Purnelle, B.; Rapoport, G.;
Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.;
Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.;
Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo,
B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.;
Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.;
Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.;
Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.;
Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.;
Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto,
K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.;
Yoshikawa, H.; Danchin, A.

#journal     Nature (1997) 390:249-256
#title       The complete genome sequence of the Gram-positive bacterium
              Bacillus subtilis.
#cross-references MUID:98044033
#accession   B70003
#status      preliminary; nucleic acid sequence not shown;
              translation not shown

##molecule_type DNA
##residues    1-272 #label KUN
##cross-references GB:299119; GB:AL009126; NID:g2635411;
              PIDN:CAB14951.1; PID:e1185846; PID:g2635457
##experimental_source strain 168

GENETICS
#gene         ytxD
CLASSIFICATION #superfamily flagellar motor rotation protein
KEYWORDS      transmembrane protein
SUMMARY       #length 272 #molecular-weight 30143 #checksum 4312

Query Match   15.4%; Score 86; DB 2; Length 272;
Best Local Similarity 34.6%; Pred. No. 6.50e+00;
Matches 18; Conservative 12; Mismatches 20; Indels 2; Gaps 1;

Db 164 IGTLVGLVLMKLNLDPMHGNMAIALTTLYGSLANVMFNPFAKLEEK 215
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1 MGSGLPLVLLTL-LGSSHGTPGWTQLQ--KLKESFLTNSSYESSFLELEK 50

RESULT 12
ENTRY   H64082      #type complete
TITLE   hypothetical protein H10630 - Haemophilus influenzae (strain
              Rd KW20)
ORGANISM #formal name Haemophilus influenzae
DATE     18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change
        24-Sep-1999
ACCESSIONS H64082
REFERENCE  A64000
#authors   Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.;
              Kirkness, E.F.; Kerlavage, A.R.; Bult, C.J.; Tomb, J.F.;
              Dougherty, B.A.; Merrick, J.M.; McKenney, K.; Sutton, G.;
              FitzHugh, W.; Fields, C.; Gocayne, J.D.; Scott, J.;
              Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman,
              J.F.; Phillips, C.A.; Spriggs, T.; Hedblom, E.; Cotton,
              M.D.; Utterback, T.R.; Hanna, M.C.; Nguyen, D.T.; Saudek,
              D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann,

```



```

TITLE      probable tyrosine kinase - alcelaphine herpesvirus 1
ORGANISM   #formal_name alcelaphine herpesvirus 1
DATE       #date_1999 #sequence_revision 24-Mar-1999 #text_change
          07-May-1999
ACCESSIONS T03130
REFERENCE   Z14840
#authors   Ensser, A.; Pflanz, R.; Fleckenstein, B.
#journal    J. Virol. (1997) 71:6517-6525
#title      Primary structure of the alcelaphine herpesvirus 1 genome.
#cross-references MIMD:97404659
#accession  T03130
#status     preliminary; translated from GB/EMBL/DDBJ
#molecule_type DNA
#residues   1-454 #label ENS
#cross-references EMBL:AF005370; NID:g2337967; PID:g2337998
SUMMARY    #length 454 #molecule_weight 50928 #checksum 4031

Query Match      15.1% Score 84; DB 2; Length 454;
Best Local Similarity 38.6%; Pred. No. 1.lle+01;
Matches 22; Conservative 11; Mismatches 20; Indels 4; Gaps 4;

Db 300 STLQVLL-VLARVMQPNSLYESWRE-IEDSS-DAYFLLLPKVLVTLHS 353
| | | | | : : : : : : : : : : : : : : : : : : : : : : : :
QY 3 SGLPLVLLLTGSGHGTGMPTQLKLKEFLTNSYSSEFFELLE-KLCLELHP 58

RESULT 15
ENTRY     S76731 #type complete
TITLE     hypothetical protein - Synchocystis sp. (strain PCC 6803)
ORGANISM  #formal_name Synchocystis sp.
#variety   PC 6803
DATE       25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
          21-Aug-1998
ACCESSIONS S76731
REFERENCE   S74322
#authors    Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
            Nakamura, I.; Miyajima, N.; Hirose, M.; Sugita, M.;
            Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
            Muraki, A.; Nakazaki, I.; Naruo, K.; Okumura, S.; Shimpo,
            S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
            Yasuda, M.; Tabata, S.
#journal    DNA Res. (1996) 3:109-136
#title      Sequence analysis of the genome of the unicellular
            cyanobacterium Synchocystis sp. PCC6803. II. Sequence
            determination of the entire genome and assignment of
            potential protein-coding regions.
#cross-references MIMD:97061201
#accession  S76731
#status     preliminary
#molecule_type DNA
#residues   1-509 #label KAN
#cross-references EMBL:D90916; GB:AB001339; NID:g1653715; PID:d1019376;
            PID:g1653732
#note       the nucleotide sequence was submitted to the EMBL Data
            Library, June 1996
SUMMARY    #length 509 #molecule_weight 56720 #checksum 9647

Query Match      15.1% Score 84; DB 2; Length 509;
Best Local Similarity 40.0%; Pred. No. 1.lle+01;
Matches 14; Conservative 7; Mismatches 12; Indels 2; Gaps 2;

Db 402 GGGLPHVYLQDLITGEANQAENYIQTLE-YLA 434
| | | | | : : : : : : : : : : : : : : : : : : : : : : : :
QY 2 GGGUPLVLLTLTGSGHGTGMPTQLKLKEFLTNSYSSEFFELLE-KLCLELHP 36

Search completed: Mon Mar 20 12:32:35 2000
Job time : 14 secs.
```

Search completed: Mon Mar 20 12:32:35 2000
Job time : 14 secs.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Mon Mar 20 12:31:17 2000; MasPar time 5.65 Seconds
Tabular output not generated.
412.201 Million cell updates/sec

Release 3.1A John F. Collins, Bioinformatics Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect score: 558
Sequence: 1 MGSGLPLVLLTLGSSHGT.....SQTSTLHARSQHVVYCVNT 78

Scoring table: PAM 150
Gap 11
Searched: 82229 seqs, 29864866 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot38
1:swissprot
Statistics: Mean 40.128; Variance 78.060; scale 0.514
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	101	18.1	501	1	SECD_AQUAE	2.16e-02
2	92	16.5	712	1	EBZE_YEAST	3.73e-01
3	91	16.3	252	1	OLP1_LYCES	5.07e-01
4	89	15.9	3005	1	2FH2_DROME	9.30e-01
5	88	15.8	1822	1	YV68_CABEL	1.26e+00
6	87	15.6	396	1	RF1M_SCHPO	1.69e+00
7	86	15.4	372	1	YTXD_BAGSU	2.27e+00
8	85	15.2	315	1	RSEB_HAEIN	3.04e+00
9	84	15.1	191	1	YDB3_SCHPO	4.07e+00
10	83	14.9	401	1	TCR2_ECOLI	5.43e+00
11	83	14.9	445	1	YGC5_ECOLI	5.43e+00
12	83	14.9	647	1	BGAL_MOUSE	5.43e+00
13	83	14.9	1233	1	NME3_HUMAN	5.43e+00
14	82	14.7	169	1	C560_HUMAN	5.43e+00
15	82	14.7	372	1	INVE_SALTY	7.22e+00
16	82	14.7	385	1	CYB_ASPNG	7.22e+00
17	81	14.5	253	1	MOTA_RHOSH	9.58e+00
18	81	14.5	261	1	ASPX_MOUSE	9.58e+00
19	81	14.5	346	1	GP41_HUMAN	9.58e+00
20	81	14.5	346	1	GP42_HUMAN	9.58e+00
21	81	14.5	386	1	CYB_SARGL	9.58e+00
22	81	14.5	387	1	CYB_PODAN	9.58e+00
23	81	14.5	482	1	LBP_RABIT	9.58e+00

24	81	14.5	5255	1	BACA_BACLI	9.58e+00
25	80	14.3	169	1	C560_BOVIN	1.27e+01
26	80	14.3	191	1	Y064_TREPA	1.27e+01
27	80	14.3	387	1	CYB_EMENI	1.27e+01
28	80	14.3	400	1	TCR8_PASMU	1.27e+01
29	80	14.3	402	1	OPDE_PSEAE	1.27e+01
30	80	14.3	411	1	VGLM_HSVBC	1.27e+01
31	80	14.3	488	1	MB1_EMENI	1.27e+01
32	80	14.3	713	1	CDGT_BACSP	1.27e+01
33	80	14.3	772	1	TF1L_HUMAN	1.27e+01
34	79	14.2	259	1	ATP6_YEAST	1.67e+01
35	79	14.2	393	1	CYB_VENIN	1.67e+01
36	79	14.2	408	1	BTN1_YEAST	1.67e+01
37	79	14.2	413	1	RF1M_YEAST	1.67e+01
38	79	14.2	493	1	ACHE_MOUSE	1.67e+01
39	79	14.2	505	1	ACHB_BOVIN	1.67e+01
40	79	14.2	511	1	MYIN_ECOLI	1.67e+01
41	79	14.2	662	1	GARP_HUMAN	1.67e+01
42	79	14.2	1015	1	PTPX_HUMAN	1.67e+01
43	78	14.0	101	1	GRO_CRIGR	2.20e+01
44	78	14.0	267	1	YTXD_BACME	2.20e+01
45	78	14.0	1239	1	NME3_MOUSE	2.20e+01

ALIGNMENTS

RESULT 1
ID SECD_AQUAE STANDARD; PRT; 501 AA.
AC O67102;
DT 15-DEC-1999 (Rel. 39, Created)
DT 15-DEC-1999 (Rel. 39, Last sequence update)
DE PROTEIN-EXPORT MEMBRANE PROTEIN SECD.
GN SECD OR AQ_973.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE; 98196666.
RA DECKERT G., WARREN P.V., GASTERLAND T., YOUNG W.G., LENOX A.L., GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R., FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus";
RL Nature 392:353-358(1998).
CC -!- FUNCTION: INVOLVED IN PROTEIN EXPORT (BY SIMILARITY).
CC -!- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
DR EMBL; AAC007060.1;
KW Protein transport; Translocation; Membrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 339 359 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 470 490 POTENTIAL.
SQ SEQUENCE 501 AA; 55459 MW; E67F690C CRC32;
Query Match 18.1%; Score 101; DB 1; Length 501;
Best Local Similarity 38.3%; Pred. No. 2.16e-02;
Matches 18; Conservative 10; Mismatches 18; Indels 1; Gaps 1;

01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-DEC-1999 (Rel. 39, Last annotation update)
OSMOTIN-LIKE PROTEIN PRECURSOR.
Lycopersicon esculentum (tomato).
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta; Eudicotyledons; Magnoliophyta; Eudicotyledons; Core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
[1]
SEQUENCE FROM N.A.
RP STRAIN-CV. VF36;
RX MEDLINE; 97128324.
CHEN R., WANG F., SMITH A.G.;
"A flower-specific gene encoding an osmotin-like protein from
Lycopersicon esculentum.";
Gene 179:301-302(1996).
[2]
SEQUENCE OF 25-33.
RP ROBERTSON D., MITCHELL G.P., GILROY J.S., GERRISH C., BOLWELL G.P.,
RA SLABAS A.R.;
Submitted (JAN-1997) to the SWISS-PROT data bank.
-1- SUBCELLULAR LOCATION: CELL WALL PROTEIN (POTENTIAL).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L76632; AAB41124.1; -;
DR HSPP; P25871; IADN
DR PROSITE; PS00316; THAUMATIN; 1.
DR PFAM; PF00314; thaumatn; 1.
KW Cell wall; Signal. 24
FT SIGNAL 1
FT CHAIN 25 252 OSMOTIN-LIKE PROTEIN.
SQ SEQUENCE 252 AA; 27265 MW; 229EB542 CRC32;
Query Match 16.3%; Score 91; DB 1; Length 252;
Best Local Similarity 54.2%; Pred. No. 5,07e-01;
Matches 13; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
Db 10 LPLSLFTLLSLQSSTNPFITL 33
||| ||||| :||| | |
QY 5 LPLVLLLLGSSRGHGPGWTLQL 28
||| ||||| :||| | |
RESULT 4
ID ZFH2_DROME STANDARD; PRT; 3005 AA.
AC P28167;
DT 01-OCT-1994 (Rel. 30; Created)
DT 01-OCT-1994 (Rel. 30; Last sequence update)
DE 01-OCT-1994 (Rel. 30; Last annotation update)
DD ZINC-FINGER PROTEIN 2 (ZINC-FINGER HOMEODOMAIN PROTEIN 2).
GN ZFH-2.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RP [1]
RP SEQUENCE FROM N.A.
RP MEDLINE; 92001539.
RX FORTINI M.E., LAI Z., RUBIN G.M.;
"The Drosophila zfh-1 and zfh-2 genes encode novel proteins
containing both zinc-finger and homeodomain motifs.";
Mech. Dev. 34:113-122(1991).
RL
CC -1- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL
CC NERVOUS SYSTEM.
CC
CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC
CC -1- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.


```

Db      164 ICTVGLGVLMKLNPHMLGNPMIAALLTTLVGSLIANWNPNTAAKLEEK 215
        ::::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     1 MGSGLPVLILLTLTGSSHGTPGMQL--KLKEFLTNSSYEVSFFLLEKK 50

RESULT    8
ID RSEB_HAEIN STANDARD; PRT; 315 AA.
AC P44792;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE SIGMA-E FACTOR REGULATORY PROTEIN RSEB HOMOLOG PRECURSOR.
OS RSEB OR HI0630.
GC Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
[1] SEQUENCE FROM N.A.
RP STRAIN-RD / KW20;
RC MEDLINE; 95305030.
RX KERNALAVAGE A.R., BULT C.J., TOMB J.-F., DOUGHERTY B.A., MERRICKS
RA MCKENNEY K., SUTTON G., FITZHUGH W., FIELDS C.A., GOCAVNE J.D.,
RA SCOTT J.F., SHIRLEY R., LIU L.-I., GLODER A., KELLEY J.M.,
RA WEIDMAN D.F., PHILLIPS C.A., SPRIGGS T., HEDBLOM E., COTTON M.
RA UTTERBACK T.R., HANNA M.C., NGUYEN D.T., SAUDER D.M., BRANDON
RA FINE L.D., FRITCHMAN J.L., FUHRMANN J.L., GEOGHAGEN N.S.M.,
RA GNEMCH I.K., McDONALD L.A., SMALL K.V., FRASER C.M., SMITH H.O.
RA VENTER J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.";
RL Science 269:496-512(1995).
SC -!- FUNCTION: SEEKS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).
CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).
CC -!- SIMILARITY: TO E.COLI RSEB AND P.AERUGINOSA MUCB
(CBY SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a c
between the Swiss Institute of Bioinformatics and the EMBL
the European Bioinformatics institute. There are no restrict
use by non-profit institutions as long as its content is
modified and this statement is not removed. Usage by and fo
entities requires a license agreement (See http://www.isb-sib.
or send an email to licensee@isb-sib.ch).

EMBL; U32746; AAC22290.1; -.
DR TIGR; HI0630; .
KW Periplasmic; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 315 SIGMA-E FACTOR REGULATORY PROTEIN
SQ FT HOMOLOG.
SQ SEQUENCE 315 AA; 35906 MW; 80BCICLD CRC32;
Query Match 15.2%; Score 85; DB 1; Length 315;
Best Local Similarity 29.5%; Pred. No. 3,04e+00;
Matches 13; Conservative 18; Mismatches 11; Indels 2;

Db      9 TALSSLULLSTASTAEELSAKQSLED-KMQQA-IDNLNVEIFAIVQ 50
        ::::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     3 SGPLPVLLLTGSGSHGTGPMTQLKLKESFLTNSYSSEFFLE 46

RESULT    9
ID YDB3_SCHPO STANDARD; PRT; 191 AA..
AC ID01356;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 21.1 KD PROTEIN C22EI2.03C IN CHROMOSOME I.
OS Schizosaccharomyces pombe (fission yeast).
GC Ascomycota; Archaeasomycetes.
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes.
```

CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-972;
 CC DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAN M.A., WALSH S.V.;
 CC Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE TH1J / PFPI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; V00611; CA23880.1; -;
 CC DR EMBL; J01830; AAS59094.1; -;
 CC DR PIR; A03507; YTECT0.
 CC DR PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
 CC KW Antibiotic resistance; Transmembrane; Inner membrane; Transport;
 CC Symptom; Transposable element.
 CC FT TRANSMEM 6 26 POTENTIAL.
 CC FT TRANSMEM 42 62 POTENTIAL.
 CC FT TRANSMEM 74 94 POTENTIAL.
 CC FT TRANSMEM 102 122 POTENTIAL.
 CC FT TRANSMEM 131 151 POTENTIAL.
 CC FT TRANSMEM 159 179 POTENTIAL.
 CC FT TRANSMEM 214 234 POTENTIAL.
 CC FT TRANSMEM 244 265 POTENTIAL.
 CC FT TRANSMEM 277 297 POTENTIAL.
 CC FT TRANSMEM 298 318 POTENTIAL.
 CC FT TRANSMEM 335 356 POTENTIAL.
 CC FT TRANSMEM 363 383 POTENTIAL.
 CC FT MUTAGEN 65 65 S->C: ALMOST NO CHANGE IN ACTIVITY.
 CC FT MUTAGEN 65 65 S->A: NO CHANGE IN ACTIVITY.
 CC FT MUTAGEN 66 66 D->N: UNABLE TO EXTRUDE TETRACYCLINE.
 CC FT MUTAGEN 66 66 D->E: MODERATE RESISTANCE TO
 CC TETRACYCLINE.
 CC FT MUTAGEN 257 257 H->E: NO H+ TRANSLOCATION.
 CC FT MUTAGEN 257 257 H->D: NO H+ TRANSLOCATION.
 CC FT CONFLICT 281 281 G -> E (IN REF. 2).
 CC FT CONFLICT 301 301 V -> D (IN REF. 2).
 CC FT CONFLICT 330 330 Q -> E (IN REF. 2).
 CC FT CONFLICT 354 354 A -> T (IN REF. 2).
 CC SQ SEQUENCE 401 AA; 43267 MW; 4823C395 CRC32;
 CC
 CC Query Match 14.9%; Score 83; DB 1; Length 401;
 CC Best Local Similarity 31.0%; Pred. No. 5.43e+00;
 CC Matches 18; Conservative 13; Mismatches 23; Indels 4; Gaps 4;
 CC
 CC Db 1 MNSSTKALVITLLDAM-GIGLIMPVLTLLRE-FIASEDIANHEGVLLA-LYALMOV 55
 CC QY 1 MGSLPVLVLLTLTSSGHTGPGMT-LQLKRESFLTNSYESSFLELLEKLCUHL 57
 CC
 CC RESULT 11
 CC ID YGCS_ECOLI STANDARD; PRT; 445 AA.
 CC AC Q46909;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 01-NOV-1997 (Rel. 35, Last annotation update)
 CC DE HYPOTHETICAL METABOLITE TRANSPORT PROTEIN IN CYSJ-ENO INTERGENIC
 CC DE REGION.
 CC GN YGCS.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC OC Escherichia.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN-K12 / MG1655;
 CC RX MEDLINE; 97426617.
 CC RA BLATTNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 CC RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 CC
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN-972;
 CC DEVLIN K., CHURCHER C.M., BARRELL B.G., RAJANDREAN M.A., WALSH S.V.;
 CC Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE TH1J / PFPI FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 270043; CAA93890.1; -;
 CC DR EMBL; 270043; CAA93890.1; -;
 CC KW Hypothetical protein.
 CC SQ SEQUENCE 191 AA; 21078 MW; D47EB47B CRC32;
 CC
 CC Query Match 15.1%; Score 84; DB 1; Length 191;
 CC Best Local Similarity 33.3%; Pred. No. 4.07e+00;
 CC Matches 14; Conservative 11; Mismatches 16; Indels 1; Gaps 1;
 CC
 CC Db 148 PVLVEENLI-TSQPGTAMLFGLKLEQVASKDKYNAVYKSL 188
 CC QY 6 PLVLLTLTGSSHTGPGMTLQLKLESFLTNSYESSFLEL 47
 CC
 CC RESULT 10
 CC ID TCR2_ECOLI STANDARD; PRT; 401 AA.
 CC AC P02980;
 CC DT 21-JUL-1986 (Rel. 01, Created)
 CC DT 21-JUL-1986 (Rel. 01, Last sequence update)
 CC DT 15-DEC-1998 (Rel. 37, Last annotation update)
 CC DE TETRACYCLINE RESISTANCE PROTEIN, CLASS B (TETA(B)) (METAL-
 CC TETRACYCLINE/H+ ANTIPOINTER).
 CC GN TETA.
 CC OS Escherichia coli.
 CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC OC Escherichia.
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE; 84109550.
 CC RA NGUYEN T.T., POSTLE K., BERTRAND K.P.;
 CC "Sequence homology between the tetracycline-resistance determinants
 CC of Tn10 and pBR322";
 CC Gene 25:83-92(1983).
 CC [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE; 83143319.
 CC RA HILLEN W., SCHOLIMIER K.;
 CC "Nucleotide sequence of the Tn10 encoded tetracycline resistance
 CC gene";
 CC Nucleic Acids Res. 11:525-539(1983).
 CC [3]
 CC RP MUTAGENESIS OF HIS-257.
 CC RX MEDLINE; 91177844.
 CC RA YAMAGUCHI A., ADACHI K., AKASAKA T., ONO N., SAWAI T.;
 CC "Metal-tetracycline/H+ antiporter of Escherichia coli encoded by a
 CC transposon Tn10. Histidine 257 plays an essential role in H+
 CC translocation";
 CC J. Biol. Chem. 266:6045-6051(1991).
 CC [4]
 CC RP MUTAGENESIS OF 65-66.
 CC RX MEDLINE; 90368755.
 CC RA YAMAGUCHI A., ONO N., AKASAKA T., NOUMI T., SAWAI T.;
 CC "Metal-tetracycline/H+ antiporter of Escherichia coli encoded by a
 CC transposon, Tn10. The role of the conserved dipeptide, Ser65-Asp66,
 CC in tetracycline transport";
 CC J. Biol. Chem. 265:15525-15530(1990).
 CC -1- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE

RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U29579; AAA69281.1; ALT_INIT.
 DR EMBL; AE000360; AAC75813.1; ALT_INIT.
 DR ECOGENE; EG13126; YGCS.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; 1.
 DR PFM; PFM0083; sugar_tr; 1.
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane.
 FT TRANSMEM 23 43
 FT TRANSMEM 57 77 POTENTIAL.
 FT TRANSMEM 86 106 POTENTIAL.
 FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 176 196 POTENTIAL.
 FT TRANSMEM 224 274 POTENTIAL.
 FT TRANSMEM 287 307 POTENTIAL.
 FT TRANSMEM 312 332 POTENTIAL.
 FT TRANSMEM 338 358 POTENTIAL.
 FT TRANSMEM 370 390 POTENTIAL.
 FT TRANSMEM 401 421 POTENTIAL.
 SQ SEQUENCE 445 AA; 48234 MW; BDD078EF CRC32;

Query Match 14.9%; Score 83; DB 1; Length 445;
 Best Local Similarity 38.6%; Pred. No. 5.43e+00;
 Matches 17; Conservative 9; Mismatches 16; Indels 2; Gaps 2;
 Db 296 GALLGLVLT-LLAHRKFLGSLIAATLVNACLPFGSSLTLL 338
 QY 23 GMTQLKLESFLTNSYE-SGFLELEKLCGLLLHLPSTSVTL 65

RESULT 12
 ID BGAL_MOUSE STANDARD; PRT; 647 AA.
 AC P23780;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE BETA-GALACTOSIDASE PRECURSOR (EC 3.2.1.23) (LACTASE) (ACID BETA-
 DE GALACTOSIDASE).
 GN GLB1 OR GLB-1 OR BGL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 91076843.
 RA NAKBA E., SUZUKI K.;
 RT "Molecular cloning of mouse acid beta-galactosidase cDNA: sequence,
 RT expression of catalytic activity and comparison with the human
 RT enzyme.";
 RL Biochem. Biophys. Res. Commun. 173:141-148(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DBA/2J;
 RX MEDLINE; 91298941.
 RA NAKBA E., SUZUKI K.;
 RT "Organization of the mouse acid beta-galactosidase gene.";

RL Biochem. Biophys. Res. Commun. 178:158-164(1991).
 CC -1- FUNCTION: CLEAVES BETA-LINKED TERMINAL GALACTOSYL RESIDUES FROM
 CC GANGLIOSIDES, GLYCOPROTEINS, AND GLYCOSAMINOGLYCAN.
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
 CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL.
 CC -1- SIMILARITY: BELONGS TO FAMILY 35 OF GLYCOSYL HYDROLASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M57734; AAA37293.1;
 DR EMBL; M75122; AAA37292.1;
 DR EMBL; M75137; AAA37292.1; JOINED.
 DR EMBL; M75107; AAA37292.1; JOINED.
 DR EMBL; M75108; AAA37292.1; JOINED.
 DR EMBL; M75109; AAA37292.1; JOINED.
 DR EMBL; M75111; AAA37292.1; JOINED.
 DR EMBL; M75112; AAA37292.1; JOINED.
 DR EMBL; M75113; AAA37292.1; JOINED.
 DR EMBL; M75114; AAA37292.1; JOINED.
 DR EMBL; M75115; AAA37292.1; JOINED.
 DR EMBL; M75116; AAA37292.1; JOINED.
 DR EMBL; M75117; AAA37292.1; JOINED.
 DR EMBL; M75118; AAA37292.1; JOINED.
 DR EMBL; M75119; AAA37292.1; JOINED.
 DR EMBL; M75120; AAA37292.1; JOINED.
 DR EMBL; M75121; AAA37292.1; JOINED.
 DR PIR; A37086; A37086.
 DR MGD; MGI:88151; BGL.
 DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
 DR PFM; PFM01301; Glyco_hydro_35; 1.
 KW Hydrolase; Glycosidase; Lysosome; Signal; Glycoprotein.
 FT SIGNAL 1 24
 FT PROPEP 25 29
 FT CHAIN 30 647 BETA-GALACTOSIDASE.
 FT ACT_SITE 189 189 PROTON DONOR (POTENTIAL).
 FT ACT_SITE 269 269 NUCLEOPHILE (POTENTIAL).
 FT CARBOHYD 27 27 POTENTIAL.
 FT CARBOHYD 248 248 POTENTIAL.
 FT CARBOHYD 500 500 POTENTIAL.
 FT CARBOHYD 504 504 POTENTIAL.
 FT CARBOHYD 510 510 POTENTIAL.
 FT CARBOHYD 544 544 POTENTIAL.
 FT CARBOHYD 557 557 POTENTIAL.
 FT CARBOHYD 617 617 POTENTIAL.
 FT CONFLICT 517 517 N -> D (IN REF. 2).
 FT CONFLICT 539 539 G -> R (IN REF. 2).
 SQ SEQUENCE 647 AA; 73121 MW; 15BCF158 CRC32;

Query Match 14.9%; Score 83; DB 1; Length 647;
 Best Local Similarity 37.8%; Pred. No. 5.43e+00;
 Matches 14; Conservative 8; Mismatches 12; Indels 3; Gaps 3;
 Db 10 LPLALLQLLGAHGIYNTORTFKLDYSDRFLKDG 46
 QY 5 LPLVLLTLTGSSHGT-G-PGWTQLKL-KESFLTNS 38

RESULT 13
 ID NME3_HUMAN STANDARD; PRT; 1233 AA.
 AC Q14957;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE GLUTAMATE [NMDA] RECEPTOR SUBUNIT EPSILON 3 PRECURSOR (N-METHYL
 DE D-ASPARTATE RECEPTOR SUBTYPE 2C) (NR2C) (NMDAR2C).
 GN GRIN2C.

Search completed: Mon Mar 20 12:31:26 2000
Job time : 9 secs.

MPRLH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Mon Mar 20 12:31:44 2000; MasPar time 14.52 Seconds
Tabular output not generated. 372.517 Million cell updates/sec

Title: >US-09-092-296-15
Description: (1-78) from US09092296.pep
Perfect Score: 558
Sequence: 1 MGSGLPLVLLTLGGSHGT.....SGTSVTLHARSQHVVYVCNT 78

Scoring table: PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 04
Listing first 45 summaries

Database: sptemb112
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phase 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 38.987; Variance 81.279; scale 0.480

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	101	18.1	501	2	PROTEIN EXPORT MEMBRAN	8.06e-02
2	93	16.7	303	11	SBF PROTEIN PRECURSOR.	9.09e-01
3	93	16.7	382	8	APOCYTOCHROME B (EC 1.	9.09e-01
4	92	16.5	209	2	RESTRICTION ENDONUCLEA	1.22e+00
5	92	16.5	863	14	ENV POLYPROTEIN.	1.22e+00
6	92	16.5	1275	13	TRANSCRIPTION CO-REPR	1.22e+00
7	89	15.9	879	2	FIL-ORF879 PROTEIN.	2.92e+00
8	88	15.8	297	8	CYTOCHROME B (FRAGMENT	3.89e+00
9	87	15.6	143	14	HERPESVIRUS TYPE 6 DNA	5.17e+00
10	87	15.6	303	11	SBF PROTEIN PRECURSOR.	5.17e+00
11	86	15.4	322	4	W08D2.3 PROTEIN.	6.86e+00
12	86	15.4	562	5	CYTOCHROME B (FRAGMENT	9.08e+00
13	85	15.2	297	8	CYTOCHROME B (FRAGMENT	9.08e+00
14	85	15.2	297	8	CHALCONE SYNTHASE HOMO	9.08e+00
15	85	15.2	395	10	SERPIN PRECURSOR.	9.08e+00
16	85	15.2	448	13	PURINERGIC P2Y11 RECEP	1.20e+01
17	84	15.1	371	4	D1H2934	1.20e+01
18	84	15.1	441	2	DIHYDROOROTASE.	1.20e+01
19	84	15.1	454	14	PUTATIVE TYROSINE KINA	1.20e+01
20	84	15.1	509	2	HYPOTHETICAL 56.7 KD P	1.20e+01

21	84	15.1	886	4	Q9Y5Q9	1.20e+01
22	84	15.1	1016	5	O17485	1.20e+01
23	84	15.1	1016	5	O17484	1.20e+01
24	83	14.9	190	14	Q9WGB5	1.58e+01
25	83	14.9	244	8	Q35487	1.58e+01
26	83	14.9	297	8	O03290	1.58e+01
27	83	14.9	344	8	Q9XPF3	1.58e+01
28	83	14.9	378	2	O66624	1.58e+01
29	83	14.9	385	8	O63316	1.58e+01
30	83	14.9	386	8	Q92440	1.58e+01
31	83	14.9	444	2	Q9XBA7	1.58e+01
32	83	14.9	711	5	O45825	1.58e+01
33	83	14.9	1236	4	O15398	2.08e+01
34	82	14.7	135	4	O75609	2.08e+01
35	82	14.7	372	2	O56052	2.08e+01
36	82	14.7	372	2	O54037	2.08e+01
37	82	14.7	372	2	O54031	2.08e+01
38	82	14.7	372	2	O54026	2.08e+01
39	82	14.7	372	2	O54027	2.08e+01
40	82	14.7	372	2	O54036	2.08e+01
41	82	14.7	372	2	O54038	2.08e+01
42	82	14.7	372	2	O54030	2.08e+01
43	82	14.7	372	2	O54028	2.08e+01
44	82	14.7	946	3	O13284	2.08e+01
45	82	14.7	1282	11	O60520	2.08e+01

ALIGNMENTS

RESULT 1						
ID	O67102	PRELIMINARY;	PRT;	501 AA.		
AC	O67102;					
DT	01-AUG-1998 (Tremblrel. 07, Created)					
DT	01-AUG-1998 (Tremblrel. 07, Last sequence update)					
DT	01-NOV-1998 (Tremblrel. 08, Last annotation update)					
DE	PROTEIN EXPORT MEMBRANE PROTEIN SECD.					
GN	SECD.					
OS	Aquifex aeolicus.					
OC	Bacteria; Aquificales; Aquificaceae; Aquifex.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-VF5;					
RX	MEDLINE; 98196666.					
RA	DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,					
RA	GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUYAY M., HUBER R.,					
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;					
RT	"The complete genome of the hyperthermophilic bacterium Aquifex					
RT	aeolicus."					
RL	Nature 392:353-358(1998).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-VF5;					
RA	DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,					
RA	GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUYAY M., HUBER R.,					
RA	FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;					
RL	Submitted (JUL-1997) to the ENBL/GenBank/DBSJ databases.					
DR	ENBL; AE000716; AAC07060.1;					
SQ	SEQUENCE 501 AA; 55459 MW; E67F690C CRC32;					
	Query Match 18.1%; Score 101; DB 2; Length 501;					
	Best Local Similarity 38.3%; Pred. No. 8.06e-02;					
	Matches 16; Conservative 10; Mismatches 18; Indels 1; Gaps 1;					
DB	454 VILDFQF-GSGPVKGFTATLGTATIASFISNVYAKVFLDLNLSKIL 499					
	: :					
Oy	8 VLLLTLLGSSHGTPGNTLQKLKESFLTNSVSSFFLELEKLCLL 54					
	: :					
RESULT 2						
ID	Q9Z0J6	PRELIMINARY;	PRT;	303 AA.		
AC	Q9Z0J6;					
DT	01-MAY-1999 (Tremblrel. 10, Created)					
DT	01-MAY-1999 (Tremblrel. 10, Last sequence update)					

```

01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE SBP PROTEIN PRECURSOR.
GN SBF.
OS Ekturis norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
RN Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
[1]
RA BOETTNER FROM N.A.
RP ROETTER M., LAEFF M., SUTER-CRAZZOLARA C.;
RT "Identification of a novel member of the TGFbeta superfamily."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ011969; CAA09891.1; -
DR EMBL; AJ011970; CAA09891.1; JOINED.
DR HSP; P18075; I BMP.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 303 AA; 33438 MW; 0IC35FCC CRC32;

Query Match 16.7%; Score 93; DB 11; Length 303;
Best Local Similarity 36.7%; Pred. No. 9.09e-01;
Matches 18; Conservative 11; Mismatches 18; Indels 2; Gaps 2;

Db 20 LFLLLLLLLSWPSQDGLALPEQRSLSESQNLNDELGRFDLLSRL 68
| : | | | | | : | : | : | : | : | : | : | : | : | : |
Qy 5 LPVLVLTLLIG-SHGTPGWTQLK-LKESFTNSSVESFELEKLK 51
RESULT 3
ID Q37395 PRELIMINARY; PRT; 382 AA.
AC Q37395;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DI 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DL 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE APOCYTOCHROME B (EC 1.10.2.2) (UBIQUINOL--CYTOCHROME C REDUCTASE)
(GYTOCHROME BC1 COMPLEX).
DN COB.
OG Allomyces macrogynus.
OS Mitochondrion.
OC Eukaryota; Fungi; Chytridiomycota; Chytridiales; Blastocladiaceae;
RN [1] Blastocladiaceae; Allomyces.
RA PAQUIN B., LANG B.F.;
RP MEDLINE; 96226032.
RT "The mitochondrial DNA of Allomyces macrogynus: the complete genomic
sequence from an ancestral fungus.";
RJ J. Mol. Biol. 255:688-701(1996).
[2]
SEQUENCE FROM N.A.
PAQUIN B., LAFOREST M.J., LANG B.F.;
Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
-!- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q + 2
FERROCYTOCHROME C.
EMBL; U41288; AAC4921.1; -
PFAM; PF00032; cytochrome_b_c_1.
PFAM; PF00033; cytochrome_b_n_1.
Oxidoreductase; Mitochondrion..
SQ SEQUENCE 382 AA; 43467 MW; A2IEIA1 CRC32;

Query Match 16.7%; Score 93; DB 8; Length 382;
Best Local Similarity 33.3%; Pred. No. 9.09e-01;
Matches 17; Conservative 15; Mismatches 16; Indels 3; Gaps 3;

Db 8 PVLSANSLDFPLSNITLYNWFGSLG-LCLVIOTGVTLAHYAPS 57
| : | | | : | : | : | : | : | : | : | : | : | : | : |
Qy 22 PGMTQLK-LAESFL-TNSSESSFELEKLCILHLPISGTVTLHHAS 70
RESULT 4
ID O68567 PRELIMINARY; PRT; 209 AA.
AC O68567;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DL 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

```

AC Q9W6S7;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE TRANSCRIPTION CO-REPRESSOR SIN3.
 GN SIN3.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
 OC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
 OC Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA VERMAAK D., WADE P.A., JONES P.L., SHI Y.-B., WOLFFE A.P.;
 RF "Functional analysis of the SIN3-histone deacetylase: RPD3-RbAp48-
 RT histone H4 connection in the Xenopus oocyte."
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF154112; AAD34644.1; -;
 SQ SEQUENCE 1275 AA; 145088 MW; EB1942F4 CRC32;
 Query Match 16.5%; Score 92; DB 13; Length 1275;
 Best Local Similarity 27.5%; Pred. No. 1.22e+00;
 Matches 11; Conservative 13; Mismatches 13; Indels 1; Gaps 1;
 Db 935 VGLKRDNDGSAQLKLEKPEMDIADYDNPFLDMVRNL 974
 QY 13 LLGSSGTGPGMTLQKLKESF-LTNSSYESSFLELEKL 51
 RESULT 7
 ID P75377 PRELIMINARY; PRT; 879 AA.
 AC P75377;
 DT 01-FEB-1997 (TrEMBLrel. 02, Created)
 DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE Fil-ORF879 PROTEIN.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M129;
 RA MEDLINE; 97105885.
 RA HIMMELREICH R., HILBERT H., PIAGENS H., PIRKL E., LI B.C.,
 RA HERRMANN R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae."
 RL Nucleic Acids Res. 24:4420-4449(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M129;
 RA HIMMELREICH R., HILBERT H., LI B.-C.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE000042; AAB96079.1; -;
 SQ SEQUENCE 879 AA; 101086 MW; 94D217E0 CRC32;
 Query Match 15.9%; Score 89; DB 2; Length 879;
 Best Local Similarity 28.9%; Pred. No. 2.92e+00;
 Matches 11; Conservative 15; Mismatches 11; Indels 1; Gaps 1;
 Db 677 EGIPKDSNY-SSFVHLDDQKSLFLQAKVSGIDINENK 713
 QY 32 ESFTNSSYESSFLELEKLCILLHLPSGTSVTLHAR 69
 RESULT 8
 ID O03305 PRELIMINARY; PRT; 297 AA.
 AC O03305;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
 DE CYTOCHROME B (FRAGMENT).
 OS Carettochelys insculpta (pitted-shelled turtle).
 OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Testudines;
 OC Cryptodira; Trionychoidae; Carettochelyidae; Carettochelys.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
 RL Syst. Biol. 0:0-0(0).
 CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C - Q + 2
 CC -1- FERRICYTOCHROME C. WHICH ARE NOT COVALENTLY
 CC BOUND TO THE PROTEIN (BY SIMILARITY).
 DR EMBL; U81355; AAB57646.1; -;
 DR PFAM; PF00032; cytochrome_b_c; 1.
 DR PFAM; PF00033; cytochrome_b_n; 1.
 KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
 KW Heme.
 FT NON_TER 1 297
 FT NON_TER 297 297
 SQ SEQUENCE 297 AA; 33587 MW; 7D5808B5 CRC32;
 Query Match 15.8%; Score 88; DB 8; Length 297;
 Best Local Similarity 31.5%; Pred. No. 3.89e+00;
 Matches 17; Conservative 10; Mismatches 25; Indels 2; Gaps 2;
 Db 165 GLTIVHLFLFYETGNNPTGLNSMD-KIPHPHYFYKDFV-GLILMLAILNL 216
 QY 4 GLPIVLLTLTGSSHGTPGMTLQKLKESFTLNSSYESSFLELEKLCILLHL 57
 RESULT 9
 ID Q69582 PRELIMINARY; PRT; 143 AA.
 AC Q69582;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
 DE HERPESVIRUS TYPE 6 DNA.
 OS Human herpesvirus 6.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94181269.
 RA THOMPSON J., CHOUDHURY S., KASHANCHI F., DONIGER J., BERNEMAN Z.,
 RA FRENKEL N., ROSENTHAL L.J.;
 RT "A transforming fragment within the direct repeat region of human
 RT herpesvirus type 6 that transactivates HIV-1."
 RL Oncogene 9:1167-1175(1994).
 DR EMBL; X73675; CAA52028.1; -;
 SQ SEQUENCE 143 AA; 13317 MW; 597857A6 CRC32;
 Query Match 15.6%; Score 87; DB 14; Length 143;
 Best Local Similarity 47.2%; Pred. No. 5.17e+00;
 Matches 17; Conservative 4; Mismatches 14; Indels 1; Gaps 1;
 Db 88 LGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLGLAGGFL 123
 QY 1 MGSGPLVLLTL-LGSSHGTPGNTLQKLKESFL 35
 RESULT 10
 ID Q9Z007 PRELIMINARY; PRT; 303 AA.
 AC Q9Z007;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
 DE SBF PROTEIN PRECURSOR.
 GN SBF OR GDF15.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-129SVJ;
 RA BOETTNER M., LAAFF M., SUTER-CRAZZOLARA C.;

```

RT "Identification of a novel member of the TGF-beta superfamily.";
RN Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RA HSIAO E.C., KONIARIAS L.G., ZIMMERS T.A., SEBALD S.M., SITZMANN J.V.,
RA HUYNH T.V., LEE S.-J.;
RT "Growth/differentiation factor-15: a new TGF-beta family member
RT induced following liver and bile duct injury.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ011967; CA098930.1;
DR EMBL: AJ011968; CA098930.1; JOINED.
DR EMBL: AF159571; RAD1410.1;
DR HSSP: I18075; I18P.
KW Signal.
FT SIGNAL
SQ SEQUENCE 303 AA; 33256 MW; FBBF45FE CRC32;

Query Match 15.6%; Score 87; DB 11; Length 303;
Best Local Similarity 36.7%; Pred. No. 5.17e+00;
Matches 18; Conservative 10; Mismatches 19; Indels 2; Gaps 2;

Db 20 LFLLLLLLWSPSGDALAMPQRRSGPESQNLNDELGRFQDLRLSRL 68
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 5 LPLVLLTLTG-SHGTGPGMTLQKL-KESFLTNSYSSFFLELEKL 51

RESULT 11
ID Q14968 PRELIMINARY; PRT; 322 AA.
AC Q14968;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE RHODOPSIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92250505.
RA KUNZ D., GERARD N.P., GERARD C.;
RT "The human leukocyte platelet-activating factor receptor. cDNA
RT cloning, cell surface expression, and construction of a novel epitope-
RT bearing analog.";
RL J. Biol. Chem. 267:9101-9106(1992).
DR EMBL: M76676; AAB97766.1;
DR PFAM: PF00001; 7tm_1; 2.
KW GTP-binding.
SQ SEQUENCE 322 AA; 33096 MW; 70F54EC8 CRC32;

Query Match 15.4%; Score 86; DB 4; Length 322;
Best Local Similarity 34.5%; Pred. No. 6.86e+00;
Matches 19; Conservative 13; Mismatches 22; Indels 1; Gaps 1;

Db 95 ALVLLIFLLSLGNCVGMVIVKHQLRTVYNFILS-LSLSDLLTALCLPAA 148
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 6 PLVLLTLTGSSHGTPGNTLQKLKESFLTNSYSSFFLELEKLCULLLPSG 60

RESULT 12
ID Q23220 PRELIMINARY; PRT; 562 AA.
AC Q23220;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JAN-1999 (TrEMBLrel. 09, Last annotation update)
DE W08D2.3 PROTEIN.
GN W08D2.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA SWINBURNE J., AINSOUGH R.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.

```

```

RN SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COFFEY T., COOPER J., COULSON A.,
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
FT Nature 368:32-38(1994).
DR EMBL: Z70271; CAA94232.1;
SQ SEQUENCE 562 AA; 63331 MW; 5272E400 CRC32;

Query Match 15.4%; Score 86; DB 5; Length 562;
Best Local Similarity 31.5%; Pred. No. 6.86e+00;
Matches 23; Conservative 19; Mismatches 27; Indels 4; Gaps 4;

Db 122 IILLICIVIVGVIGLISLAVISN-FVIFK-QQSFL-LLOFLHLIGAFSGVTAMFY 178
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 7 LVLLTLTGSSHGTPGNTLQKLKESFLTNSYSSFFLELEKLCULLLHLPSTVTLH 66

Db 179 QSAPESEHOEDCNS 191
: : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 67 H-ARSOHHVWCNT 78

RESULT 13
ID Q03295 PRELIMINARY; PRT; 297 AA.
AC Q03295;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
OS Trachemys scripta (Red-eared slider turtle).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Testudines;
OC Cryptodira; Testudinoidea; Emydidae; Trachemys.
RN [1]
RP SEQUENCE FROM N.A.
RA SHAFFER H.B., MEYLAN P., MCKNIGHT M.L.;
RL Syst. Biol. 0:0-0(0).
CC -1- CATALYTIC ACTIVITY: OH(2) + 2 FERRICYTOCHROME C = Q + 2
CC -1- FERROCYTOCHROME C.
CC -1- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
DR EMBL: U81351; AAB57666.1;
DR PFAM: PF00032; cytochrome_b_c; 1.
DR PFAM: PF00033; cytochrome_b_n; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Transmembrane;
KW Heme.
FT NON_TER 1
FT NON_TER 297
SQ SEQUENCE 297 AA; 33479 MW; 1BCBCFAC CRC32;

Query Match 15.2%; Score 85; DB 8; Length 297;
Best Local Similarity 33.3%; Pred. No. 9.08e+00;
Matches 18; Conservative 9; Mismatches 25; Indels 2; Gaps 2;

Db 165 GLTLVHLFLHETGSGNNPTGLNSVD-KIPFHPYFSYKD-LLGILMLTLTL 216
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 4 GLPLVLLTLTGSSHGTPGNTLQKLKESFLTNSYSSFFLELEKLCULLLH 57

RESULT 14
ID Q03294 PRELIMINARY; PRT; 297 AA.
AC Q03294;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)

```


